True/False Questions

1. If data on (Y, X) are available at only two values of X, the models $Y = \beta_0 + \beta_1 X + \epsilon$ and $Y = \beta_1 X + \beta_2 X^2 + \epsilon$ will fit the data equally well.
2. A "$H_0 : \beta = 0$" t-test on one parameter can also be performed as an F-test, and $t^2 = F$ for the values of the test statistics.
3. If a $(1 - \alpha)$ confidence interval for slope β_1 contains zero, we will not reject $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ at the α level.
4. If we fit the model $Y = \beta_0 + \epsilon$ to a set of data, we will always get $b_0 = \overline{Y}$ and $\hat{Y} = \overline{Y}$.
5. Choice of a test level α implies that we expect our test decision to be wrong $100\alpha\%$ of the time when we reject H_0.
6. When $R^2 = 1$, all the data lie on a line of positive, negative, or zero slope.
7. In regression work we are looking for an empirical relationship between a response and one or more predictor variables.
8. A straight line $\hat{Y} = b_0 + b_1 X$ fitted by least squares must contain the point $(\overline{X}, \overline{Y})$.
9. The assumption that the errors in the model are normally distributed is not needed for the construction of the analysis of variance table, which is simply an algebraic breakup of ΣY_i^2.
10. There are always exactly as many normal equations as there are parameters in the model.
11. $\Sigma(X_i - \overline{X})(Y_i - \overline{Y}) = \Sigma X_i Y_i - n\overline{XY}$.
12. If we fit the model $Y = \beta X + \epsilon$ to a set of data (X_i, Y_i), $i = 1, 2, \ldots, n$, the sum of the residuals is not necessarily zero.
13. In regression work, the assumption that the errors in Y are normally distributed is needed to validate use of F- and t-tests.
14. Even if we are using the wrong regression model, repeat runs can be used to provide an estimate of σ^2.
15. We obtain the "lack of fit" degrees of freedom by subtracting the "pure error" degrees of freedom from the "residual" degrees of freedom.
16. The R^2 statistic is not the average of the squared correlations of Y with X_1, Y with X_2, $\ldots$, Y with X_k, where $X_1, X_2, \ldots, X_k$ are the X's being fitted in the model.
17. A correctly calculated "joint confidence region" for β_0 and β_1 (say) has an elliptical shape.
18. The confidence interval statements we make about individual β's depend on the $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$ assumption.
19. Even if we fit the wrong model, it would be possible for some or all of the parameter estimates we got to be unbiased (by the terms omitted).
20. The formula $\hat{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}s^2$ provides the estimated variances and pairwise covariances of the b's we have fitted.
21. $H_0 : \beta_1 = \beta_2\beta_3$ is not a linear hypothesis.
22. The extra $SS(\mathbf{b}_2|\mathbf{b}_1)$, obtained from fitting the model $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, is exactly the same as the regression sum of squares for the model $\mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, provided that $\mathbf{X}_2'\mathbf{X}_2 = \mathbf{0}$.
23. *Both* of the following are true for a linear model: (a) $\Sigma e_i = 0$ when there is a β_0 in the model; (b) $\Sigma e_i \hat{Y}_i = 0$ *always*.

24. Even though we assume that the errors ϵ_i in a linear model are pairwise uncorrelated, this is not typically true of the residuals e_i, $i = 1, 2, \ldots, n$.
25. The plot of e_i versus Y_i always has a slope of size $1 - R^2$ in it.
26. The Durbin–Watson statistic always lies between 0 and 4, no matter what the (nonsingular) linear regression problem may be.
27. If we have ten residuals, five positive and five negative, there are only two ways that two runs can occur out of the 252 possibilities of rearrangements of signs. (Assume 252 is right; don't worry about *that* aspect.)
28. When we add a second predictor X_2 to a model $Y = \beta_0 + \beta_1 X_1 + \epsilon$, the value of s^2 may go down, or up, but the value of R^2 cannot go down.
29. Confidence bands for the true mean of Y given X can be evaluated for a straight line model fit, and similar bands exist when there are two or more predictors.
30. A test of $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$ in the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ can be made either in (a) extra sum of squares F-test form, or (b) $t = b_2/(\text{Est. Var}\ (b_2))^{1/2}$ form, where $t = F^{1/2}$.
31. On a normal probability plot, a line should be drawn through the "middle bulk" of the residuals as a check on normality.
32. An observation can be both influential and an outlier.
33. An observation can be influential but not an outlier.
34. An observation can be an outlier but not influential.
35. If we fit a straight line $Y = \beta_0 + \beta_1 X + \epsilon$, and we find that $b_0 = 0$ exactly, then the residuals will still add to zero.
36. We can justify the use of least squares when $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$ via the application of maximum likelihood.
37. The extra sum of squares for b_1 and b_2 (say) given $b_3, b_4, \ldots, b_q$ has two degrees of freedom, no matter how many other b's are "given."
38. $F(1, 22) = \{t(22)\}^2$.
39. The specific values of the (sequential) sums of squares for a series of input variables in a regression may be changed if we change the "order of entry" of those input X's.
40. $b_w = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ is the "generalized" least squares solution where we assume $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$.
41. $(\mathbf{Y} - \hat{\mathbf{Y}})'\mathbf{1} = 0$ is true when the model contains a β_0 term.
42. The R^2 statistic is the square of the correlation between the $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ columns.
43. The rectangle formed by individual confidence intervals for β_1 and β_2 (say) is not a correct "joint region" for the pair (β_1, β_2).
44. The formula $E(\mathbf{b}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2\boldsymbol{\beta}_2$ tells us the bias effect on $\mathbf{b}$ of failing to include the terms $\mathbf{X}_2\boldsymbol{\beta}_2$ in the model.
45. If, in the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$, then the elements of $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ are also normally distributed.
46. The vector of fitted values $\hat{\mathbf{Y}}$ is always orthogonal to the vector of residuals $\mathbf{e}$.
47. When we fit the model $Y = \beta_0 + \epsilon$ (i.e., no X's) to a set of data, $R^2 = 0$, always.
48. When $R^2 = 1$, all the residuals must be zero.
49. If there are n observations and e degrees of freedom for pure error, a unique fitted linear model can contain no more than $(n - e)$ parameters, including β_0.
50. The linear hypothesis $H_0 : \beta_1 - \beta_2 = \beta_2 - \beta_3 = \beta_3 - \beta_4 = 0$ will take up three degrees of freedom in a formal test.
51. The model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2 \ln X_1 + \epsilon$ is a linear model.
52. The model $Y = \beta_0 + \beta_1 X + \beta_2 (\beta_3)^X + \epsilon$ is not a linear model.
53. If we "know" that σ_Y is proportional to the kth power of the response then we can use Y^{1-k} as a variance stabilizing transformation.

54. For binomial data, arcsin $(Y^{1/2})$ is a useful variance stabilizing transformation.

55. The model $Y = \theta + \alpha X_1X_3 + \beta X_1 + \alpha\beta X_2 + \epsilon$ with parameters (θ, α, β) is nonlinear.

56. The model $Y = \beta_0 + \beta_1(X_1 - X_2) + \beta_2(X_1 - X_2)^2 + \epsilon$ is a linear model.

57. If a model $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ is fitted and a t-test indicates we should keep the b_2X^2 term in, we should also retain the b_1X term, significant t-test or not.

58. The transformation form $V = (Y^\lambda - 1)/(\lambda\, \dot{Y}^{\lambda-1})$ for $\lambda \neq 0$ is continuous at $\lambda = 0$, where $\dot{Y}$ is exp $\{n^{-1} \Sigma_{i=1}^{n} \ln Y_i\}$.

59. Even though we choose the best value of λ in the transformation form $V = (Y^\lambda - 1)/(\lambda\, \dot{Y}^{\lambda-1})$, we may still not get a good regression fit.

60. The $\sin^{-1}(Y^{1/2})$ transformation stabilizes the variance if the original data Y_i are binomial data.

61. After a set of regression data has been "centered and scaled," the coefficients in the normal equations are *all* correlations ρ, such that $-1 \leq \rho \leq 1$.

62. In "analysis of variance" problems, we should always work out the residuals and examine them after fitting the model.

63. "Analysis of variance" models are usually overparameterized, that is, have more parameters then we can estimate uniquely.

64. To take account of level differences between two groups of observations, we can make use of a dummy variable Z such that $Z = -1$ for the first group, and $Z = 1$ for the second group.

65. Fitting the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ and fitting the model $Y = \beta_0' + \beta_1(X_1 - \overline{X}_1) + \beta_2(X_2 - \overline{X}_2) + \epsilon$, by least squares, both lead to the same fitted equation.

66. The correlation matrix of the estimated regression coefficients is a rescaled form of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix.

67. If two columns of the $\mathbf{X}$ matrix are proportional to each other [e.g., column $A = 4$ (column B)], then the determinant of $\mathbf{X}'\mathbf{X}$ will be zero.

68. Even if the determinant of $\mathbf{X}'\mathbf{X}$ is zero, it is still possible to solve the least squares normal equations $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$ but the solution will not be unique.

69. The backward elimination (selection) procedure can leave out of consideration a good combination of the X's that it "never gets to."

70. The "usual" F-statistic percentage point does not provide an accurate determination of the true percentage value at which entry or exit (removal) tests are made in selection procedures.

71. Ridge regression calculations for $\theta \neq 0$ provide residual sums of squares that are always greater than the least squares values.

72. In a certain metric, ridge regression can be regarded as "least squares subject to a restriction that the sum of squares of the parameters (except the intercept) is restricted to a spherical region."

73. If X is measured over a small range and the corresponding regression coefficient estimate b is *not* significant, X may, nevertheless, have an important effect on Y, which the data do not reveal.

74. The C_p statistic $[\text{RRS}_p/s^2 - (n - 2p)]$ will always take the value p exactly for at least one regression equation in the "all regressions" case.

75. Use of the backward elimination method may reveal ill-conditioning in the data at the first step.

76. $S_{YY} = (n - p)s^2/(1 - R^2)$ can be used to calculate S_{YY} from any regression printout, but it may be subject to some round-off error.

77. To fit a planar model uniquely in mixture ingredients X_1, X_2, X_3 such that $X_1 + X_2 + X_3 = 1$, we cannot use the model function $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ but must modify it.

78. $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$ lies in the estimation space and is always unique, even when $\mathbf{b}$ is not unique.

79. $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ lies in the error space.

80. The estimation space is spanned (defined by) the column vectors of the $\mathbf{X}$ matrix.

81. Even if the columns of $\mathbf{X}$ are linearly dependent, they still define the estimation space.

82. If $\mathbf{M} = \mathbf{X}'\mathbf{X}$ is singular, we can still find a generalized inverse $\mathbf{M}^-$ such that $\mathbf{M}\mathbf{M}^-\mathbf{M} = \mathbf{M}$.

83. A generalized inverse is not unique.

84. If $\mathbf{Y}$ is a vector and $\mathbf{P}$ is a projection matrix, $\mathbf{PY} = \mathbf{P}^2\mathbf{Y} = \cdots = \mathbf{P}^m\mathbf{Y}$.

85. The fact that $\mathbf{PY} = \mathbf{P}^2\mathbf{Y} = \mathbf{P}(\mathbf{PY})$ means that, after the $\mathbf{Y}$ vector has been projected onto $R(\mathbf{P})$, projecting it again does not change it, as it is already in $R(\mathbf{P})$.

86. If $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the projection matrix for a regression problem with model containing a β_0, the rows and columns of $\mathbf{P}$ all add to 1.

87. If $\mathbf{X}$ is not of full rank, we can still estimate some linear *combinations* of the β's uniquely, but not all of the individual β's.

88. A linear combination $\mathbf{c}'\boldsymbol{\beta}$ of the parameters can *always* be estimated provided $\mathbf{c}'$ is expressible as a linear combination of the rows of $\mathbf{X}$, and even if $\mathbf{X}$ is not of full rank.

89. Analysis of variance models are typically overparameterized, and some choice of suitable dummy variables must be made for a nonsingular regression treatment.

90. Residuals should be examined in analysis of variance type problems as well as in general regression problems.

91. "Analysis of variance" data analyses are really just variations of regression problems.

92. A one-way classification typically consists of several groups or "treatments" with multiple observations in each group.

93. The one-way classification model $Y_{ij} = \mu + t_i + \epsilon_{ij}$ must have a restriction attached to it and this restriction is not unique.

94. To begin doing a nonlinear estimation problem some "starting values" or "initial values" are needed.

95. Linear least squares approximations are used to get approximate confidence intervals for the parameters in nonlinear estimation problems.

96. $Y = \beta_1 \exp\{\beta_2 t\} + \epsilon$ is a nonlinear model.

97. The least squares method is valid for estimating the parameters of a nonlinear regression model.

98. Iterative procedures are needed to solve a general nonlinear regression problem.

99. The sum of squares surface for a nonlinear model is not an ellipsoidal "bowl" in general.

100. The model $Y = \exp(\theta_1 + \theta_2 t) + \epsilon$ is nonlinear in θ_1 and θ_2.

101. Fitting the model $\ln Y = \theta_1 + \theta_2 t + \epsilon$ would usually provide reasonable starting values for fitting the model $Y = \exp(\theta_1 + \theta_2 t) + \epsilon$.