|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Week 5 | Feb 14  - log-linear tends to fix non-linearity, non-constant variance, and right skew in response variable  - linear-log tends to fix non-lin + right skew in x vars  - Log-log combines both “fixes” - If symmetric then mean=median - Unequal variance: spread in residuals in. to the right - variance increases with the predicted values - log-linear slope: when x increases by 1 unit, predict the median of Y to change by a factor of eb1  - linear-log slope: when x increases by p%, we predict Y to change by b­1ln(1+p/100)  - log-log slope: when x increases by p%, we predict the median of Y to change by a factor of (1+p/100)b1   |  |  |  | | --- | --- | --- | |  | x | Ln(x) | | Y | Linear: Y^=b0+b1x | Lin-log: Y^=b0+b1ln(x) | | Ln(y) | Log-lin:  Ln(y)^=b0+b1x | Log-log  Ln(y)^=b0+b1ln(x) |   - suppose you increase by 10%, p=0.10 (1.10b1) | Feb 16  - R^2=69%: 69% of the variability in happiness can be explained by a linear model including life expectancy. - Residual=y - y^, (pos residual = underpredicted)  - regression line for sample: y^ = b0 + b1x  - regression line for population: µ = β0 + β1x  - MSE quantifies variability of regression model  - SSE-sum of squared error, variance=σ^2  - CI bands are narrower b/c it’s estimating a population mean + distribution of means has less variability than individuals like in PI  - CI: we are 95% confident that the population mean \_y\_ is btwn \_\_ and \_\_ for \_\_ with a ­x  - PI: 95% confident that the \_\_ is btwn \_\_ and \_\_ with a x |
| Week 6/7 | Feb 23  - Ho:β1=0; Ha:β1>0-positive,β1<0-negative,β1≠0-noclr - Assume approx normal/mean of β1 and SD of σb1 = σ/sqrt(SSxx) - When don't know σ, replace w/ RMSE=σhat=sqrtMSE - error: We expect about 95% of the observed values y to be within 2x(RMSE) of their predicted values ŷ - SSxx(sum) = # obs, variability in xvalues; > var = >precise estimate of slope - slope quality values 0-21>5-10,even when same #obs - We are 95% confident that when(X)inc by1,the pop. mean (y)dec between \_\_&\_\_ | Feb 28    - F & t statistic related when k=1(SLR), both test Ho:β1=0,(ANOVA always)Ha:β1≠0,F=t^2,p-value=same - Overall Method Significance: Ex. Maybe, if F (3) and t (9) are high=low p-value (<0.0001), R^2 (SStotal - SSE / SSTotal) is fairly small @ 0.122, statistically significant but many not practically significant - If large n=always get small p-value, not due to relationship, only a lot of evidence. |
| Week 7/8 | Mar 2  - slope: The mean/predicted salary for late career professors is $7,331.74 more than early career prof. - Intercept: The predicted happiness for group 1 is the (y-int) and (slope) for group 1 - Slope: The predicted happiness when alcohol use Is light (non-reference grp) is 0.72 lower (slope=-) than those who abstain from alcohol use (reference grp) - Int:When DAS score is 0 and the student does not regularly put all-nighters, predicted happiness is \_\_.  - Slope: When DAS score inc by 1, pred happiness dec by 0.158, assuming that all-nighters is held constant. - Residual:The happiness for the 1set student in the dataset was underpredicted by 0.865 - Con:There is \_\_evidence that high values of DAS scores are linearly associated w low happiness scores on average, for the pop., when all-nnighters is held constant. | Mar 7  - There is moderate evidence to suggest that there is a relationship btw all-nighter and happiness after controlling for DAS Scores - Ho:β1=β2=β3=0; Ha:At least 1 βi≠0 - CI:We are 95% confident that when DAS increases by 1, the pop. mean is different for only alcohol use category based on the FTest - We are 95% confident the pop. mean happiness who regularly pulls all-nighters is between 0.5 and 3.51 lower than those who do not do all-nighters, when DAS scores are held constant - slope:For every additional hour spent, we predict an inc. # of mailings of 13.821 for customers not awr srvc. |
| Week 8/9 | Mar 9  - additive model (no interaction terms)  - for a sample: Y^ = b0 + b1X1 + b2X2 +…+ bkXk  - for a population: µ = β0 + β1X1 + β2X2 +…+ βkXk  - 4 conditions:  - Linearity: examine residual plot (linear model is appropriate and mean of residuals is 0)  - Independence: values of yi are indep (rand sample)  - Normality: normal qq plot, residual normal quantile plot (residuals follow a normal distribution)  - Equal variability: examine residual plot (residuals have equal variability. The variance of y is constant for all values of x) | Mar 21  - interaction term: cross product of two or more x’s  - interaction: rltn btwn 1 explanatory var and the response var depends on a third var  - “there are interactions among the explanatory vars”  - interactions ≠ correlation  - interpreting slope of interaction btwn 2 quant vars: refers to how much the slope btwn Y and X1 changes for every 1 unit increase in X2  - larger diff in slopes, greater the interaction |
| Week 9/10 | Mar 23  - ex: “R2=0.12: 12% of the variability in happiness can be explained using a linear model including poor sleep quality. [after accounting/adjusting for the complexity of the model (adjusted R2 - penalty for adding new x)]”  - larger R2 correspond w/ better pred of y using x vars  - when another x var is added to model, R2 can’t decre.  - multicollinearity: when 2+ x vars are correlated  - if multicol: signs of slopes could be opposite of expected, standard errors become inflated  - to detect multicol:  - correlations btwn vars (|r| > 0.9), VIF > 5  - T-tests for all/nearly all indiv slopes not sig, but F-test for overall model sig (or R2 high)  - VIF: amount of change in SE due to correlations of that x var w/ other x vars  - sols: drop 1+ vars, new var as combo of correlated vars, more data | Mar 28  - quadratic terms if there is a quad rltnship btwn Y and X  - population/theoretical model: µ = β­0 + β­1X1 + β­2X21  - sample/estimated model: Y^ = b0 + b1X1 + b2X12  - if include quadratic term, also include first-order term  - int interpretation: “When X is 0, predict Y to be int.”  - 1st-order term gives rate of change when X=0  - slope of quadratic term: > 0 (curve up), < 0 (curve down)  - slope for distance2 is pos, indicating a concave up rltnsh  - partial F-tests: testing x vars w/ more than 2 categories, testing subsets of vars  - AlcoholUse has 4 levels so make 3 dummy vars: each dummy var will be coded as 1 if true, 0 if false. If AlcoholUseLight = 1, true. Abstain is reference group. If all dummy vars = 0, Abstain = true.  - nested model: model subset of another model |
| Week 10/11 | Mar 30  - too few variables: bad predictions because of bias  - too many variables: increase variability of predictions  - model selection: techniques to make reasonable model (include necessary vars, exclude unnecessary)  - include vars of importance, look at F-test, adjusted R2, RMSE, Mallow’s Cp. Consider multicol + assumption  - the more x vars, the large sample size needed  - overall model fit: smallest p-val from F-test, smallest RMSE, highest adjusted R2, low Mallows Cp, lowest # of vars, most # of sig vars (t-test)  - Backward Elimin: (1) fit full model, (2) find term w/ largest p-val (3) if largest p-val is small, that is your chosen model (4) if largest p-val is large, elimin that term and fit your new model (5) go back to step 2  - α=0.05 cutoff for sequential model selection | Apr 4  - (outlier) unusual points: (1) unusual vals of Y (2) unusual vals of Y given val of X (3) unusual val of X  - high leverage point: point w/ extreme val for 1 or more of X vars (outlier in the x-direction)  - high influence point: point (usually a high leverage point) that influences the regression model  - residual very far from 0: outlier for regression  - standardized residual: z= residual/RSME  - high stand residual if > 2 or < -2 [very high if > 3 or < -3]  - high lever val if h > 2((k+1)/n) [h>3((k+1)/n)] if very high  - Cook’s D combines standardized residual and leverage  - high influence if Cook’s D>0.5, very high influence if D >1 |