|  |  |  |
| --- | --- | --- |
| Week 1 | Jan 17  - variable: anything collected  - statistics: summary of data collected in sample  - parameter: summary of population  - random sample: randomly pick sample from set  - random assignment: randomly assign units different treatments  - variability: uniqueness, different (used to understand to make informed decisions leads to accuracy)  - intended population (conceptual, hypothesized): includes everyone interested in  - actual population: sampling frame, may be same as intended population  - sample: every data from, subset of actual population  - categorical variable (qualitiative): nominal; ordinal  - quantitative var. (numerical): discrete; continuous  - explanatory variable (independent, prediction)  - response variable (dependent) | Jan 19  - mean: paramater (µ), statistc (x̄)  - standard deviation: parameter (σ), statistic (s)  - proportion: parameter (p), statistic (p̂)  - regression slope: paramter (β), statistic (β^)  - model: data = model + error --> y=f(X, parameters) +ε  - research process: research question, make a plan, collect data (choose model, fit model to data, assess and use data)  - experiment data collection: assign treatment, control  - observational study collection: no randomly assigned treatments, units random, probability sample  - sample survey collection: participants rand., probability sampling  - random sampling: method choosing participants  - random assignment: treatment, randomization  - sampling bias, response bias, nonresponse bias  - Simple Rand Samples, stratified, cluster, multistage |
| Week 2 | Jan 24  - Center: mean-common, median - data not symmetric  - mean->outliers; smaller data-> greater impact  - median=50th percentile (middle #)  - mean=fair share value (balancing point)  - summaries not affected by outliers=robust (resistant)  - center: median=robust, mean=not robust  - observations in tail influence the mean  - observations (usually) don't influence median  - right skew: mean>median; left skew: mean<median  - position->sample quartiles:5#summary, IQR, boxplot  - shape=histogram; comparing=boxplot  - spread(var):IQR=Q3-Q1~range of middle 50%, robust  - s=sample, σ=pop;values closer to mean=<s,far=larger  - s≥0,s=0 when observations are equal, s^2=variance  - typically skew>uniform>bell shaped >smallest  - stand.dev (not robust): how far, on avg, the observations are from the mean | Jan 26  - Parameter=population mean= μ  - Statistic=sample mean= x̄  - Confidence Interv: point estimate+/-margin of error  - single mean: x̄ ± tα/2(s/√n)  - Se=s/√n & multiplier is a T quartile  - t distribution: bell shaped  - population: se= σ/√n; stnadard error of mean  - s: poor estimator of σ when sample sizes are small  - T: normal distribution; more variability than normal distribution especially for smaller sizes  - specified by degrees of freedom (df); df = n-1  - inc n -> dec me -> dec confidence interval  - CI: SRS, pop. Norm. distr. (n≥30), approx. norm.  - we are \_\_\_% confident that the population mean \_\_\_ size is between \_\_\_and\_\_\_ units; [ ]  - actual pop.mean where \_\_is\_\_similar ones in sample  - null->Ho, Alt->Ha(H1); 1-sided:Ha>/<, 2-sided:Ha≠ |
| Week 3 | Jan 31  - hypothesis test same conditions as CI  - "Considering the sample was not randomly selected, we are not positive on what the actual population is"  - t=(x̄-μo)/se=( x̄-μo)/(s/√n)  - p-value: prob of obs≥extreme result when Ho is true  - >0.1-little to no,>0.05-weak, >0.025-moderate, >0.001-strong,<0.001-overwhelming  - p-value:The prob of having a sample mean pH as extreme or more than\_\_IF the null hypothesis were true was\_\_, a fairly small prob  - inc n->low p-values->diff small->se is extremely small  - Paired: sample avg diff=đ, parameter=μd  - CI=đ±tα/2(sd/√n); assume norm. distr./n≥30  - hyp: Overwhelming evidence to suggest there is a difference in the pop. mean scre before and after the program  - We are \_% confident the true mean is btw & units | Feb 2  - Interpret scatterplot-form, direction, strength  - Outlier in explanatory variable (x-direction)  - Outlier in response variable (y-direction)  - Points outside general pattern: extreme outlier  - points near pattern-regression outlier  - Correlation coefficient: pop-⍴(rho), sample-r  - r-#, no units, linear association btw 2 variables, can be affected by outliers, -1≤r≤1, r=0-no linear relationship; r=-1 straight down, r=1 straight up, scale and flipping x and y doesn't affect r  - ex:the linear relationship btw\_&\_ is (strength) & (+/-)  - correlation≠causation  - lurking var=confounding, other that affect relationship btw 2 var., quant/qual  - causation=ran assn/control all var & change only exp  - Simple linear regression-variables not interchangable  - analyze: conditions, linear->scatter, vars.= (exp/resp) |
| Week 4 | Feb 7  - predicted value=ŷ, observed=y, residual value = e = error = amt of variation in y model can't acct for  - Sum of residuals=0; slope=b1  - For a 1 unit inc. in exp var. (x), the predicted value of the response variable (y) will inc/dec by amt of slope  - "", the value of the response variable (y) will inc/dec by the amt of slope, on average  - Intercept=bo; When the value of the exp var. is 0, the predicted value of the resp var (y) = to the y-int  - slope: +=above=underpredicted, -=under=overpred.  - Obs values = larger = more variability, predicted = smaller = less variability  - mean: ŷ, SD: Sŷ; Residual(e): mean: 0, SD: Se  - R^2=%, ratio:0-1: EX: \_\_% of the variability in (y)\_\_\_ can be explained using a linear model w (x)\_\_\_.  - Outlier in y=large resid, x=high leverage: >influence  - avg=stronger relationship, not accurate | Feb 9  - SLR-linear relationship between x&y w/mean error=0, examine scatter; Errors ind. rand. Samp.; errors normally distributed(QQ, histogram of residuals);errors have equal variability, residual plot, = variability.  - Residual:y-residual,x-predicted,horiz.line@residual=0  - homoscedasticity(=variability)=fan shaped  - conditions not met:curve,fan shaped  - small sample with outliers=bad  - LSR used when residuals not normally distributed as long as n is sufficiently large  - Check histogram of resid.,approx norm.  - Ind met=random sample, linearity met=scatter linear, resid. Plot no curves, homoscedasiticity met=residual plot has no curves, normal met=histogram of residuals is bell-shaped, QQ plot close to linear. |
| Week 5 | Feb 14  - log-linear:fix non-linearity, non-constant variance&right skew in response variable  - linear log:fix non-linearity and right skew in exp var.  - 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  - If both reciprocal and ln look reasonable then ln  - log-linear: (lny)^=bo+b1x; When x inc. by 1 unit, we predict the median of y to change by a factor of e^b1  - ln-log: ŷ=bo+b1ln(x); When x inc. by P%, we predict y to change by b1ln(1+P/100)  - Log-log:(lny)^=bo+b1ln(x); When x inc. by P%, we predict the median of y to change by a factor of (1+P/100)^b1 | Feb 16  Conditions:  - Linearity-met/not met. (1) Scatterplot is linear/curved (2) Residual plot curved/not curved  Independence: Met/not (random sample/ based on data, variables are independent  - Normal: Met/not. (1) Histogram is bell shaped  (2) QQ plot is linear  - Homoscedacity: Met/not. Residual plot (fan/no fan)  - R^2=69%: 69% of the variability in happiness can be explained by a linear model including life expectancy.  - Residual=y-yhat, the happiness for the US was underpredicted by 1.21 |
| Extra Space | - DESCRIBE DISTRIBUTION: skewed right/left/uniform, outliers, median, IQR = Q75%-Q25%  - STANDARD DEVIATION: "the avg diff btw the (total prices of diamonds) to the mean (price $\_\_) is (stdev)"  - "\_\_ is the explanatory variable that we will use to predict \_\_\_, the response variable. The explanatory variable is always on the X-axis and the response on the Y-axis  - CORRELATION: sqrt(r^2)  - SIMPLE LIN REGRESSION CONDITIONS: independence, randomly sampled, linearity, homoscedasticity (residual plot fan/no fan), normality  - OBSERVATIONAL STUDY: no treatments administere, no random assignment | Click or tap here to enter text. |