Statistics Study Guide

Matthew Chesnes
The London School of Economics

September 22, 2001

1 Descriptive Statistics

- Pictures of Data: Histograms, pie charts, stem and leafs plots, scatter plots, OGives
  - OGive: A measure of cummulative distribution percentages. Compute reasonable intervals for data and determine their frequency, then their cummulative frequency, and finally their percentage frequency (percentile). Plot percentages at the upper end of the intervals. Easy way to display quartiles.
  - Stem and Leaf Plot: Stem is the major part of the data and leaves are minor part. Choose a reasonable unit for the stem. Organize leaves in order of magnitude. If leaves are too large, split stem into smaller intervals. Placing the leaves equidistant, provides a histogram like representation.
  - "A Diagram is Interacting the Eye" - B. Blight.
- Measures of Data: Mean, Median, Mode, Standard Deviation, Quartiles
- Right Skewed Data - positively skewed - long right tail
- Left Skewed Data - negatively skewed - long left tail
- Measures of Location and Spread
  - Mean: Average (stable value and useful in analysis though sensitive to outliers).
  - Mode: Data value that occurs most often ... highest peak in the pdf of the continuous case.
  - Median: 50^{th} percentile (Insensitive to outliers though not as useful in statistical inference)
  - Range: Max - Min (Crude and inaccurate)
  - Interquartile Range: $75^{th} - 25^{th}$ quartile.
  - Sample standard deviation, s, an estimate of population standard deviation, σ,

$$s = \sqrt{\frac{\text{Corrected sum of Squares}}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

  - The standard deviation is calulated on n-1 degrees of freedom rather than n because dividing by n would yield a biased estimator.
  - Alternative form of s:

$$\text{CSS} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \text{Raw sum of squares - correction}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}}.$$

- Location Transformations: $y = x + 5 \longrightarrow sd(y) = sd(x)$.
- Scale Transformations: $y = 5x \longrightarrow sd(y) = 5sd(x)$.

2 Probability

- Additive Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- Exclusive Events: $P(A \cup B) = P(A) + P(B)$.
- Total Probability: $P(A) + P(A^c) = 1$.
- Demorgan's Laws: $P(A^c \cap B^c) = P((A \cup B)^c)$, and $P(A^c \cup B^c) = P((A \cap B)^c)$.
- Combinatorial: ${}^nC_x = \dfrac{n!}{x!(n-x)!}$.

2.1 Conditional Probability - Baysian Statistics

- Bayes Rule: $P(A|F) = \dfrac{P(A \cap F)}{P(F)} = \dfrac{P(F|A) * P(A)}{P(F)}$.
- Independence: A and B are independent iff: $P(A|B) = P(A)$. Therefore $\dfrac{P(A \cap B)}{P(B)} = P(A)$. Thus, $P(A \cap B) = P(A) * P(B)$. **Only if A and B are independent.**
- Exclusive events are VERY dependent. One happening completely excludes the possibility of the other occuring.
- The Law of Total Probability: $P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B) * P(B) + P(A|B^c) * P(B^c)$.
- So in general, Bayes Law can be written,

$$P(B_i|A) = \frac{P(A|B_i) * P(B_i)}{\sum_{i=1}^{n} P(A|B_i) * P(B_i)}.$$

3 Discrete Probability Distributions of Random Variables

3.1 The Binomial Distribution

- Bernoulli Trials: A series of n independent trials under the same circumstances. The probability of success, P(success), in all trials is identical.
- $P(\text{k successes in n trials}) = ({}^nC_k)p^k(1-p)^{n-k}$.
- Cumulative Distribution Function: $F_k = P(K \leq k)$.

3.2 Other Discrete Distributions

- Hypergeometric: Not independent trials, sampling without replacement. As $n \rightarrow \infty$, hypergeometric $\rightarrow$ Binomial.
- Negative Binomial: The distribution of the number of trials needed to get k successes.
- Multinomial: Generalization of binomial for more than 2 classifications.

3.3 The Poisson Distribution

- Discrete Distribution
- Applications: Measuing random arrival times: components that break down over time, Defective items in a large batch.
- Memoryless Property: A every point in time, there is always the same chance of the event occurring.
- Arrival rate: λ.
- $P(\text{r arrivals in time t}) = \dfrac{\lambda^r e^{-\lambda}}{r!}$.
- The rate, λ, is in terms of time, t.
- Use Poisson approximation for the binomial when n is large AND p is either large (~ 1) or small (~ 0). Thus, use the poisson approximation if $np < 10$. Then $PDF_{Poisson} = \dfrac{(np)^r e^{-np}}{r!}$.

4 Properties of Random Variables

- A random variable is not a number, it's a concept.
- The mean of X.
  - The probability weighted average of all the values that the random variable can take.
  - $\mu = \sum_{i=1}^{n} x_i p_i = E[X]$.
  - If X is distributed binomially, $E[X] = np$.
  - If X is distributed poisson, $E[X] = \lambda$.
  - Expectation is a linear operator.
  - $E[a + bX] = a + bE[X]$.
- Variance and Standard Deviation of X.
  - R is a random variable
  - $\sigma_R^2 = E[(R - \mu)^2] = \sum_r (r - \mu)^2 p_r$.
  - Alternate Form: $\sigma_R^2 = E[(R)^2] - (E[R])^2$.
  - Rearranging for an important and useful result: $\sigma_R^2 + \mu^2 = E[(R)^2]$.
  - If X is distributed binomially, $\sigma_X^2 = npq$.
  - If X is distributed poisson, $\sigma_X^2 = \lambda$.

5 Continuous Probability Distributions

- X is a continuous random variable.
- $P(X \leq 2) = P(X < 2)$.
- Cumulative Distribution Function: $CDF = F(x) = P(X \leq x)$.
- $F(-\infty) = 0$. and $F(+\infty) = 1$.
- $E[X] = \mu = \int_{-\infty}^{+\infty} xf(x)dx$.
- $\sigma_X^2 = E[(X)^2] - (E[X])^2$.

5.1 The Uniform Distribution

- Constant PDF over the range of the distribution.

5.2 The Exponential Distribution

- Consider the Poisson process with points occuring at random in time. λ is the average number of occurances per unit of time. The time between occurances is a continuous random variable, X, and it follows an exponential distribution.
- $1 - F(x) = P(X > x) = P($0 occurances over the interval from (0,x)$) =$

$$\frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}.$$

- Thus $F(x) = 1 - e^{-\lambda x}$.
- Thus $f(x) = \frac{d}{dx}(1 - e^{-\lambda x}) = \lambda e^{-\lambda x}$.
- If X is distributed exponentially, $E[X] = \frac{1}{\lambda}$.
- If X is distributed exponentially, $\sigma_X^2 = \frac{1}{\lambda^2}$.

5.3 The Normal Distribution

- The Central Limit Theorem: If n value are sampled from a population and if n is sufficiently large, then the sample mean (or sum) is normally distributed whatever the distribution of the parent population.
- If parent is normal, "n large" $\rightarrow$ n relatively small.
- If parent is very nonnormal, "n large" $\rightarrow$ about 50 at most.
- Standard Normal Distribution: $\mu = 0, \sigma = 1$.

- Probability Density function for standard normal:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

- The general normal probability density function $X \sim N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

- Standardization procedure: $z = \dfrac{x-\mu}{\sigma}$
- Combining Random Variables
  - $E[aX + bY] = aE[X] + bE[Y] = a\mu_X + b\mu_Y$.
  - $\sigma^2[aX + bY] = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab * Cov(X, Y)$.
  - $\sigma^2[aX - bY] = a^2\sigma_X^2 + b^2\sigma_Y^2 - 2ab * Cov(X, Y)$.
  - $\sigma^2[3 + 2X] = 4\sigma_X^2$.
  - Theorem: Any linear function of normal variables is itself normally distributed.
- Normal Approximation to the binomial: If $R \sim$ Binomially with n trials and p, the probability of success, as $n \to \infty$, but p remains constant, $R \to$ Normal. As $n \to \infty$, but np remains constant (therefore $p \to 0$), $R \to$ Poisson (Use Poisson if $np < 10$). If $R \to$ Normal, $R \sim N(np, npq)$. IMPORTANT ... when using the normal approximation to the binomial, remember to add or subtract a half when computing intervals or finding critical values to reflect the discreteness of the original distribution.

6 Sampling Theory

- Let $X \sim N(\mu, \sigma^2)$.
- Let $Q = \sum_i x_i \forall x_i \in X$.
- Then $Q \sim N(n\mu, n\sigma^2)$.
- Thus $\bar{X} = \frac{Q}{n}. \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
- The Standard Deviation of $\bar{X} = \dfrac{\sigma}{\sqrt{n}} =$ The standard error.
- Parameters, Estimators, and Standard Errors.
  - Parameter $= \mu$; Estimator $= \bar{x}$; Standard Error $= \dfrac{\sigma}{\sqrt{n}}$.
  - Parameter $= p$; Estimator $= \dfrac{r}{n}$; Standard Error $= \sqrt{\dfrac{pq}{n}}$.
  - Parameter $= \mu_X - \mu_Y$; Estimator $= \bar{x} - \bar{y}$; Standard Error $= \sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}$.
  - Parameter $= p_1 - p_2$; Estimator $= \frac{r_1}{n_1} - \frac{r_2}{n_2}$; Standard Error $= \sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$.
- Distribution of Sample Variance:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

- Application of χ^2:

$$\frac{\sum (x_i - \bar{x})^2}{\sigma^2} = \frac{CSS}{\sigma^2} = \frac{S_{xx}}{\sigma^2} \sim \chi^2_{n-1}.$$

- $E[\chi^2_n] = n$.
- $\sigma^2[\chi^2_n] = 2n$.

7 Estimation

7.1 Point Estimation

- We want to estimate some parameter, Θ using an estimator, μ.
- Calculate The Mean Square Error = MSE = $E[(\mu - \Theta)^2]$.
- Square out to find MSE = $\sigma_\mu^2 + (E[\mu] - \Theta)^2$.
- Or otherwise written, MSE = Variance + bias^2.
- Desirable properties of estimators: unbiased: E[estimator] = parameter. Efficient: Small variance.
- For example, $E[s^2] = E[\frac{CSS}{n-1}] = E[\frac{\sum(x-\bar{x})^2}{n-1}] = \sigma^2$.
- Hence dividing by the n-1 is explained because it gives us an unbiased estimator.
- However, efficiency is more important than unbiasedness. If one estimator is slightly biased but extremely efficient, use it because of the high variability of the alternative.

7.2 Interval Estimation

- A 95% confidence interval for the mean, μ, of a normally distributed X is,

$$\mu \in (\bar{x} \pm Z_{crit}(2.5\%) * SE(\bar{x}) = \mu \in (\bar{x} \pm 1.96 * \frac{\sigma}{\sqrt{n}}).$$

- An incorrect interpretation of this interval would be: "There is a 95 percent chance that $\bar{x}$ is within 1.96 standard errors of μ."
- A correct (purist) statement would be: "if you took many samples and calculated the confidence interval for a parameter each time, then 95 percent of the confidence intervals would contain the true value of the parameter."
- This is because the interval is the *thing* that has variability, not μ. μ is a constant.
- Confidence intervals for proportions:

$$p \in (\frac{r}{n} \pm Z_{crit}\sqrt{\frac{\frac{r}{n}(1-\frac{r}{n})}{n}}).$$

- A 95 percent CI for Comparing Proportions:

$$p_1 - p_2 \in (\frac{r_1}{n_1} - \frac{r_2}{n_2} \pm Z_{crit}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}).$$

- Sample size Determination. Define d to be the Tolerance or the half length of the confidence interval. To obtain a 95 percent confidence interval for a mean to be with in a certain tolerance, d, set $n = \frac{Z_{crit}\sigma}{d}^2$. One may have to estimate σ with s using a small sample first and then determine optimal n.
- In general, $d = Z_{crit} * SE$, and the SE involves n, so solve for n and plug in d.
- Exact formulation of the variance of $\bar{x}$:

$$Var(\bar{x}) = \frac{\sigma^2}{n}(1 - \frac{n}{N}).$$

If N is very large, this correction term has little to no effect.

7.3 Confidence Intervals for Small samples

- Suppose the sample is small and the variance is unknown. A confidence interval for μ is,

$$\mu \in (\bar{x} \pm t_{n-1}^{crit} * \frac{s}{\sqrt{n}}).$$

- The t distribution, AKA, the student's t distribution, is more spread out to allow for the variability of both $\bar{x}$ and s. If σ is known, use Z distribution for sure. (Unless n is incredibly low). If n is large, use Z because even though the t distribution is theoretically correct, $t \to Z$ as $n \to \infty$.
- One other case: if n is small and the distribution is really not normal (the central limit theory does not apply), then one must use a non paratmetric approximation.
- Comparison of Means: 3 cases.
  - Paired Data. Calculate $d_i = x_i - y_i$. We want an estimate for $\mu_d = \mu_x - \mu_y$. So confidence interval becomes,

$$\mu_d \in (d \pm t_{n-1}(\frac{s_d}{\sqrt{n}})).$$

    We use the t distribution because n is small and we are estimating s_d.
  - Unpaired Large Samples. $\mu_x - \mu_y$ estimated by $\bar{x} - \bar{y}$. Thus the standard error here is,

$$S_{\bar{x}-\bar{y}} = \sqrt{\frac{S_{\bar{x}}^2}{n_{\bar{x}}} + \frac{S_{\bar{y}}^2}{n_{\bar{y}}}}.$$

    And thus a confidence interval becomes,

$$\mu_x - \mu_y \in (\bar{x} - \bar{y} \pm Z_{crit} * \sqrt{\frac{S_{\bar{x}}^2}{n_{\bar{x}}} + \frac{S_{\bar{y}}^2}{n_{\bar{y}}}}).$$

– Unpaired Small samples. Must make the assumption that the variances of the two samples is the same! Risky assumption. Assume $\sigma_1 = \sigma_2 = \sigma_p$. Thus,

$$S_p = \sqrt{\frac{CSS_1 + CSS_2}{n_1 + n_2 - 2}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}.$$

And,

$$SE = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Notice that S_p^2 is a weighted function of the sample variances with each's degrees of freedom as the weights. The test statistic for a hypothesis test or a confidence interval will follow a t distribution with $n_1 + n_2 - 2$ degrees of freedom.

7.4 Confidence Intervals for a Variance

- $S^2 = \frac{S_{xx}}{n-1}$. S^2 is not normally distributed. Since $E[S^2] = \sigma^2$, we can rearrange the terms and it can be shown that, $\frac{S_{xx}}{\sigma^2} \sim \chi^2_{n-1}$. Read off Chi squared values off the table for upper and lower limits for $n - 1$ degrees of freedom. Then,

$$0.95 = P(\underline{\chi}^2 < \frac{S_{xx}}{\sigma^2} < \bar{\chi}^2).$$

$$0.95 = P(\frac{S_{xx}}{\bar{\chi}^2} < \sigma^2 < \frac{S_{xx}}{\underline{\chi}^2}).$$

Thus,

$$\sigma^2 \in (\frac{(n-1)S^2}{\bar{\chi}^2_{n-1}}, \frac{(n-1)S^2}{\underline{\chi}^2_{n-1}}).$$

8 Hypothesis Testing

- Testing H_0 versus H_1.
- Always choose the null hypothesis to be the simpler of the two alternatives.
- Type I Error: Rejecting H_0 when it is true. (α)
- Type II Error: Failing to reject H_0 when it is false. (β)
- α and β both decrease with a larger sample size.
- Power Function: The probability of accepting H_1 (rejecting H_0) for different values of the true parameter, μ.
- Some might use the terminology, "Accepting H_1." But this would be incorrect if it implies proof. All we are saying is that the available data supports the hypothesis. Purists would never just accept, they would use the terminology, "Fail to reject H_0.
- To carry out test, define hypotheses, compute test statistic and compare with the relevant distribution.
- If n is large, use the Z distribution for your decision.
- If n is smaller and σ is unknown, use the t distribution.
- If a test statistic is on a division point of the critical values, maybe you cannot confidently reject H_0, but you should be very suspicious that it is actually true.
- Always report lowest possible α level (highest possible confidence). Doing otherwise is just ignorant. - C.Dougherty
- The P value of the test tells you exactly where the test statistic lies: it's the probabilty that under the null hypothesis, you observe an estimate as or more extreme then your value.
- When computing standard errors for test, always compute them with null values. Since we are assuming that the null is true until proven guilty, one must use its values when doing the test.
- Advantage of Paired Test: must less sensitive.
- Never use the data to form your hypothesis: choose the nature of the test (one tailed or two tailed, null and alternative hypotheses, etc) first and then carry out the test using the data.

9 Tests for Association

- Association: Relating factors via cross tabulated data. (catagorized data)
- Correlation: Relating variables via measurement data.
- Display data in a mXn contingency table. Where m and n are the number of factors your comparing, not the levels. Usually just a 2X2.
- Test H_0 = No Association Versus H_1 =Association.
- After setting up the tables you have your O's (observed data).
- Compute the E's (Expected data) as, $E = \frac{\text{Row Total * Column Total}}{\text{Grand Total}}$.
- Find Peirceson's Test Statistic as,

$$P = \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(r-1)(c-1)}.$$

- The larger the statistic, P, the larger the likelihood of rejecting H_0 in favor of Association.
- The statistic is distributed as a Chi Squared with (row-1)(col-1) degress of freedom.

10 Further Properties of Random Variables

- Let R be a random variable with p.d.f, p_r.
- Let T be some function, $\phi(R)$.
- Then Prob(T=t) = $\sum p_r$ where $\sum$ is over all the values of r such that $\phi(r) = t$. Work out the distributions of R and then T to see that this is true.
- Theorem: For a random variable X and a random variable $Y = \phi(X)$ such that ϕ is a monotonic function, the c.d.f. for X equals the c.d.f. for Y. $F(x) = G(y)$.
- Also, (IMPORTANT THEOREM), for the same transformation, ϕ, $g(y) = f(x)|\frac{dx}{dy}|$.
- For a general transformation on a random variable (ϕ not necessarily monotonic), just look at the graph of the transformed X, and evaluate the above theorem over each monotonic section.
- Joint density functions of two random variables: $f(x, y)$. This is simply a surface in three dimensions with the volume under the surface (instead of area under the curve) representing probability. Total volume under the surface is again equal to one. All of Baye's calculus on probabilities also applies to density functions.
- $f(y) = \int f(y|x)f(x)dx$.
- If X and Y are independent, $f(x, y) = f(x)f(y)$.

10.1 Covariance and Correlation

- Covariance: $Cov(X, Y) = \gamma = E[(X - \mu_X)(Y - \mu_Y)]$.
- If $\gamma > 0$, X and Y work in the same direction.
- If $\gamma < 0$, X and Y work in the opposite direction.
- It can also be shown that $Cov(X, Y) = E[XY] - E[X]E[Y]$.
- Since the covariance depends on the units of the random variable, we define the correlation coefficient to be,

$$\rho = \frac{\gamma}{\sigma_X \sigma_Y}.$$

- ρ is the "Linear Correlation Coefficient," and it lies between -1 and 1.
- If X and Y are independent, it can be shown that the $Cov(X, Y) = 0$
- If X is a linear function of Y, then $\rho_{XY} = \pm 1$.
- Properties of Variance and Covariance.

- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$.
- Variance is a "second-order" operator.
- Variances always add, though the covariance term takes the sign of ab.
- 3 variable case: $Var(aX + bY + cZ) = a^2Var(X) + b^2Var(Y) + +c^2Var(Z) + 2abCov(X, Y) + 2acCov(X, Z) + 2bcCov(Y, Z)$.
- $Cov(aX+bY, cS+dT) = acCov(X, S)+adCov(X, T)+bcCov(Y, S)+bdCov(Y, T)$.

11 Matrix Notation for the Multivariate Normal Distribution

- Define $\vec{X}$ to be a p column vector of random variables.
- $\vec{X} \sim N(\vec{\mu}, \vec{\sum})$.
- $\vec{\sum}$ is the correlation matrix. All diagonal elements of this matrix are the variances of each of the random variables. The off diagonal entries are covariances. It is of course a symmetric matrix.
- $\vec{\sum} = E[(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^T]$.
- Theorem: If $\vec{X} \sim N(0, \vec{\sum})$, (ie, is multivariate normal), Then $\vec{X}^T \vec{\sum}^{-1} \vec{X} \sim \chi_p^2$.

12 Correlation and Regression

- 6 Basic Statistics needed for regression.
- n, sample size; $\bar{x}$, sample mean of independent variable; $\bar{y}$, sample mean of dependent variable; S_{xx}, the corrected sum of squares for the x's; S_{yy}, the corrected sum of squares for the y's; S_{xy}, the corrected sum of products for x and y.
- $S_{xx} = \sum_i (x_i - \bar{x})^2$.
- $S_{yy} = \sum_i (y_i - \bar{y})^2$.
- $S_{xy} = \sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2 = \sum_i (x_i y_i) - n\bar{x}\bar{y}$.

12.1 Correlation Coefficient and Tests

- Covariance $= c = \dfrac{S_{xy}}{n-1}$.
- Correlation $= r = \dfrac{c}{S_x S_y} = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$.
- A correlation of zero means that there is no "linear" relationship between X and Y but does not necessarily mean there is no relationship at all ... it could be nonlinear.
- Test for Correlation: $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$.
- Test Statistic $= \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

12.2 Simple Linear Regression

- Use scatterplots of the data as a starting point.
- Simple Linear Model: $y_i = \alpha + \beta x_i + \epsilon_i$.
- Error term, $\epsilon_i \sim_{iid} N(0, \sigma^2)$.
- Least Squares Criteria: Minimize wrt α and β,

$$Q = \sum_i (y_i - \alpha - \beta x_i)^2.$$

- Yields estimators,

$$b = \frac{S_{xy}}{S_{xx}}.$$
$$a = \bar{y} - b\bar{x}.$$

- It can be shown that a and b are B.L.U.E. : Best, Linear, Unbiased, Estimators.

- Property: $(\bar{x}, \bar{y})$ always lies on the LS regression line.
- Property: $\sum_i e_i = 0$. (The residuals always sum to 0).
- In general, with n observations and p parameters (including the constant),

$$s^2 = \frac{\sum_i e_i^2}{n-p} \text{is an unbiased estimator of} \sigma^2.$$

- Define RSS = Residual Sums of Squares = $\sum_i (y_i - a - bx_i)^2$. Thus RSS = $\sum_i e_i^2$.
- Thus $s^2 = \dfrac{RSS}{n-p}$.
- Arranging the definition of RSS, we find,

$$RSS = S_{yy} - b^2 S_{xx}.$$

  Or in other words, RSS is the extra variability in y that we cannot explain after fitting the model.
- If S_{yy} is the total variability, then $b^2 S_{xx}$ is the explained variability.
- Analysis of Variance Table.

Source	Degrees of Freedom	Sums of Squares	Mean Square
Regression (Explained)	$p-1$	$b^2 S_{xx}$	S_r^2
Residual (Unexplained)	$n-p$	$RSS = S_{yy} - b^2 S_{xx}$	$S^2 = \frac{RSS}{n-p}$
Total	$n-1$	S_{yy}	S_T^2

- Define,

$$R^2 = \frac{b^2 S_{xx}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} S_{yy}}.$$

- R^2 is the percentage of the variability in y that is explained by the independent variables via the regression equation. In words, it is the explained variability over the total variability, so it is a good measure of how well the line fits the data.
- In simple linear regression, we saw the that the correlation coefficient, $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$. Thus, in SLR, $R^2 = r^2$. This doesn't apply to multiple regression because there we have many correlations and only one R^2 value.
- Adjusted R^2. Good for comparing models in the multiple regression setting. Reflects the addition of more variables, while always increasing R^2, might lead to a worse model. Define,

$$R^2_{adj} = \frac{S_T^2 - S^2}{S_T^2}.$$

- Standard errors for inference purposes:

$$SE(b) = \frac{\sigma}{\sqrt{S_{xx}}}.$$

$$SE(a) = \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}.$$

- Hypothesis tests and inferences about α and β are the same as always and will follow a t distribution because we are estimating σ.

- F Test for Regression: particularly useful for multiple regression. In SLR, $F = t^2$. Test $H_0 : B_i = 0 \; \forall \; i$ versus $H_1 : \beta_i \neq 0$ for at least one i. The Null hypothesis is that the regression has no effect. Test statistic $F = \frac{S_r^2}{S^2}$. If F is much different from 1, then reject H_0 and conclude that there is a valid regression effect. It can be shown that as a ratio of two Chi squared variables,

$$\frac{S_r^2}{S^2} \sim F_{p-1,n-p}.$$

12.3 Prediction Intervals

- Plug your x value into the regression equation and get your predicted y. Be careful though of points outside the range of your data. For an interval of confidence, develop a prediction interval for y. $\hat{y}$ is your estimator and the standard error of y is,

$$SE(\hat{y}) = \sigma\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}.$$

- So your prediction interval becomes,

$$y \in (\hat{y} \pm t_{n-2} SE).$$

- From the last term in the SE formula, it is clear that the further away from the mean you are, the larger your prediction interval.

12.4 Multiple Regression

- Model: $\vec{y} = \vec{x}\vec{\beta} + \vec{\epsilon}$. Where,

$$\vec{x} = \begin{bmatrix} 1 & x_{11} & x_{12} & ... & x_{1p} \\ 1 & x_{21} & x_{22} & ... & x_{2p} \\ 1 & x_{31} & x_{32} & ... & x_{3p} \\ 1 & x_{41} & x_{42} & ... & x_{4p} \\ ... & ... & ... & ... & ... \\ 1 & x_{n1} & x_{n2} & ... & x_{np} \end{bmatrix} \tag{1}$$

And,

$$\vec{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ ... \\ \beta_p \end{bmatrix}. \tag{2}$$

- Thus, for OLS, we minimize:

$$(\vec{y} - \vec{x}\vec{\beta})^T(\vec{y} - \vec{x}\vec{\beta}).$$

Which yeilds,

$$\hat{\vec{\beta}} = b = (\vec{x}^T\vec{x})^{-1}(\vec{x}^T\vec{y}).$$

13 Time Series

- Index numbers for a series of prices, $p_0, p_1, ..., p_t, ...$
- Index $= \hat{p}_t = 100\frac{p_t}{p_0}$.
- Laspeyres Index: For comparing prices using quantity at the base time, t_0.

$$P_t = 100 * \frac{\sum q_0 p_t}{\sum q_0 p_0}.$$

- Paasche Price Index:

$$P_t = 100 * \frac{\sum q_t p_t}{\sum q_t p_0}.$$

- Quantity Index:

$$P_t = 100 * \frac{\sum q_t p_0}{\sum q_0 p_0}.$$

- Value Index:

$$P_t = 100 * \frac{\sum q_t p_t}{\sum q_0 p_0}.$$

- Index Linking. Useful to reindex from time to time, but to avoid jumps, define,

$$P_t = P_{0,t} \text{ for } t = 0, ..., 10.$$

$$P_t = \frac{P_{10,t} - P_{0,t}}{100} = p_{0,10} * \frac{\sum q_{10} p_t}{\sum q_{10} p_{10}} \text{ for } t = 10, ..., 20.$$

- Time Series: $x_0, x_1, x_2, ..., x_t, ...$
- Classical Economic Time series: $x_t = T_t + S_t + C_t + I_t$. (Trend + Seasonal + Cyclical + Irregular stationary component.)
- Stationary Time Series: Relate variable to itself using 1 or more lags.
- Autoregression: $(x_t - \bar{x}) = b(x_{t-1} - \bar{x})$.
- Auto Correlation (τ is the number of lags) :

$$r_\tau = \frac{\frac{\sum (x_t - \bar{x})(x_t + \tau - \bar{x})}{n - \tau - 1}}{\frac{\sum (x_t - \bar{x})^2}{n-1}}.$$

- Models
  - 1^{st} order Autoregressive Model.

$$x_t = \lambda x_{t-1} + \epsilon_t.$$

  Where, $\lambda \in (0, 1)$ for a stationary time series and $\lambda > 1$ for a non-stationary time series.

- 2^{nd} order Autoregressive Model.

$$x_t = \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \epsilon_t.$$

- Moving Average Model.

$$x_t = \epsilon_t + b\epsilon_{t-1}.$$

$$x_{t+1} = \epsilon_{t+1} + b\epsilon_t.$$

$$x_{t+2} = \epsilon_{t+2} + b\epsilon_{t+1}.$$

  Where every neighboring term is correlated with each other, but others are not. This can be extended to more than one lagged interaction.

- Mixed Models : ARMA Model - AutoRegressive Moving Average Models -

$$\sum_{i=0}^{p} a_i x_{t-i} = \sum_{i=0}^{q} b_i \epsilon_{t-i}.$$