Formulas

ECON 601 Notes – Dr. Schreyer

Descriptive Statistics

- Number of observations n
- The arithmetic mean of x: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
- The standard deviation of x: $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
- The variance of x: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
- Covariance between x and y: $cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$
- Correlation between x and y: $r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}} = \frac{cov(x,y)}{s_x s_y}$

Random Variable Fundamentals

- Expected value of random variable X: $E(X) = \mu$
- Variance of random variable X: $var(X) = E[(X-\mu)^2] = \sigma^2$
- Standard deviation of random variable X: $\sqrt{var(X)} = \sigma$
- Standardize $X \sim N(\mu, \sigma^2)$: $Z = \frac{X-\mu}{\sigma}$

Inferences for a Population Mean

- Level of significance α

For a random variable with a known variance: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

- Two-sided hypothesis test[1]: Compare $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ with critical value $Z_{\alpha/2}^c$.
- Confidence interval: $\bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

For a random variable with an unknown variance: $\bar{X} \sim N(\mu, ?)$

- Two-sided hypothesis test: Compare $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ with critical value $t_{\alpha/2,n-1}^c$
- Confidence interval: $\bar{x} \pm t_{\alpha/2,n-1} \times \frac{s}{\sqrt{n}}$

---

[1] The notation $z_{\alpha/2}$ refers to the value of Z such that the probability to the right of this value equals $\alpha/2$.

Simple Linear Regression

- Population Model: $y_i = \beta_0 + \beta_1 x_i + e_i$ where e_i is assumed $N(0, \sigma_e^2)$ and independent.
- OLS slope: $b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$
- OLS intercept: $b_0 = \bar{y} - b_1 \bar{x}$
- Standard error of b_1: $s_{b_1} = \sqrt{var(b_1)} = \sqrt{\frac{\sigma_e^2}{\sum(x_i - \bar{x})^2}}$ where σ_e^2 is estimated with $s_e^2 = MSE$
- Standard error of b_1: $s_{b_0} = \sqrt{var(b_0)} = \sqrt{\sigma_e^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right)}$

Multiple Linear Regression

- Population Model: $y_i = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \beta_3 x_{3_i} + \ldots + \beta_k x_{k_i} + e_i$
- Number of independent variables k

Inferences for $\boldsymbol{\beta}$

- Two-sided hypothesis test: Compare $t = \frac{b - \beta^*}{s_b}$ with critical value $t^C_{\alpha/2, n-k-1}$
- Confidence interval: $b \pm t_{\alpha/2, n-k-1} \times s_b$

Sum of Squares

- Total sum of squares: $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ with $n - 1$ degrees of freedom
- Error/residual sum of squares: $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ with $n - k - 1$ deg. of freedom
- Regression/model sum of squares: $SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ with k degrees of freedom
- Relationship: $SST = SSE + SSR$
- Mean square, total: $MST = \frac{SST}{n-1} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$
- Mean squared error: $s_e^2 = MSE = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}$
- Mean square of the regression: $MSR = \frac{SSR}{k} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{k}$

Assessing the Fit

- F statistic: Compare $F = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/(n-k-1)}$ with critical value $F^C_{\alpha\,;\,k,\ n-k-1}$
- R squared: $1 - \frac{SSE}{SST} = \frac{SSR}{SST}$ = coefficient of determination
- Root MSE: $s_e = \sqrt{MSE} = \sqrt{\frac{SSE}{n-k-1}}$ = standard error of the regression

Comparing the Fit with a Nested Model

- Adjusted R squared: $1-\frac{SSE/(n-k-1)}{SST/(n-1)}=1-\frac{MSE}{MST}$
- Partial F test: Compare $F=\frac{\left(\frac{SSE_{\text{reduced}}-SSE_{\text{full}}}{(k-l)}\right)}{\left(\frac{SSE_{\text{full}}}{(n-k-1)}\right)}$ with critical value $F^{C}_{\alpha\,;k-l,\ n-k-1}$, where

  the full model has k indep. vars. and the reduced model has l indep. vars.

Predictions / Forecasts for a Simple Linear Regression[2]

- Conditioning value: x_m
- Point prediction: $\hat{y}=b_0+b_1x_m$
- Prediction interval: $\hat{y}\pm t_{\frac{\alpha}{2},n-k-1}\times s_p$ where

  $$s_p=\sqrt{s_e^2\left(1+\frac{1}{n}+\frac{(x_m-\bar{x})^2}{(n-1)s_x^2}\right)}=\text{standard error of the forecast}$$

- Confidence interval: $\hat{y}\pm t_{\frac{\alpha}{2},n-k-1}\times s_m$ where

  $$s_m=\sqrt{s_e^2\left(\frac{1}{n}+\frac{(x_m-\bar{x})^2}{(n-1)s_x^2}\right)}=\text{standard error of the prediction}$$

- Relationship: $s_e^2=s_p^2-s_m^2$
- Mean square deviation: $MSD=\frac{\sum_{i=1}^{n2}(y_i-\hat{y}_i)^2}{n2}$ where

  $n2$ is the number of observations in the hold-out sample

- Root mean squared error:[3] $RMSE=\sqrt{MSD}$

Multiple Linear Regression (using matrix algebra)

- Population Model: $Y=X\beta+e$ where

  Y is a $n\times 1$ vector of observations;

  X is a $n\times k$ matrix of n observations on k independent variables;

  β is a $k\times 1$ vector of population coefficients;

  e is a $n\times 1$ vector of errors;

  and assume $e_i\sim N(0,\sigma_e^2)$ and are independent.

- OLS coefficients: $b=(X'X)^{-1}X'Y$ where

  b is a $k\times 1$ vector of estimated coefficients.

- Variance-covariance matrix[4] $\sigma_e^2(X'X)^{-1}$ where σ_e^2 is estimated with $s_e^2=\frac{e'e}{n-k}$

---

[2] The case of a multiple regression model is excluded due to the complexity of the formulas for s_p and s_m.

[3] The term *root mean squared error* is the same term that can be used to describe the *standard error of the regression* $=s_e=\sqrt{MSE}$ when assessing the fit of an estimated regression. Although they are different formulas, they share the same basic idea of examining the square root of the average squared mistake.

[4] The standard errors of the OLS coefficients are given by the square root of the main diagonal of the variance-covariance matrix.

The Standard Normal Distribution
(Probabilities between 0 and the value of z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

The t Distribution

Degrees of freedom	Area in Upper Tail				
	0.10	**0.05**	**0.025**	**0.01**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576