COMM 581 – Final Exam Formulae & Distribution Tables

Simple & Multiple Linear Regression

Sums of squares

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$$

$$SSreg = \sum_{i=1}^{n} \left(\bar{y} - \hat{y}_i\right)^2 = b_1 SPxy$$ (Second formula is for SLR only)

$$SSy = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^{\,2} - \left(\sum_{i=1}^{n} y_i\right)^2 \Big/ n = s_y^{\,2}(n-1)$$

SSy = SSreg + SSE

Test statistics

$$F = \frac{SSreg / m}{SSE / (n-m-1)} = \frac{MSreg}{MSE}$$

F critical = $F_{m,\, n-m-1,\, 1-\alpha}$

$$t = \frac{b_j - c}{s_{b_j}}$$

$t_{critical} = t_{n-m-1,\, 1-\alpha/2}$

Measures of goodness of fit

$$r^2 = \frac{SSy - SSE}{SSy} = 1 - \frac{SSE}{SSy} = \frac{SSreg}{SSy}$$ = R^2 (for multiple linear regression)

Pseudo-R^2 = I^2 = 1 – SSE / SSY (in original units)

$$R_a^{\,2} = 1 - \left(\frac{n-1}{n-(m+1)}\right)\frac{SSE}{SSy}$$

$$SE_E = \sqrt{\frac{SSE}{n-m-1}}$$

Simple Linear Regression

Co-efficients and standard errors

$b_0 = \bar{y} - b_1 \bar{x}$

$$b_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}(n-1)}{s_x^{\ 2}(n-1)} = \frac{SPxy}{SSx}$$

$$s_{b_0} = \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{SSx}\right)} = \sqrt{\frac{MSE \times \sum_{i=1}^{n} x_i^{\ 2}}{n \times SSx}}$$

$$s_{b_1} = \sqrt{\frac{MSE}{SSx}}$$

Confidence Intervals for co-efficients

For β_o: $b_0 \pm t_{1-\alpha/2,n-2} \times s_{b_0}$

For β_1: $b_1 \pm t_{1-\alpha/2,n-2} \times s_{b_1}$

Confidence interval for mean of y given x_h

$$\hat{y} \mid x_h \pm t_{n-2,1-\alpha/2} \times s_{\hat{y}|x_h}$$

$$\hat{y} \mid x_h = b_0 + b_1 x_h$$

$$s_{\hat{y}|x_h} = \sqrt{MSE\left(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{SSx}\right)}$$

Prediction Interval

$$\hat{y}_{(new)} \mid x_h \pm t_{n-2,1-\alpha/2} \times s_{\hat{y}(new)|x_h}$$

$$\hat{y}_{(new)} \mid x_h = b_0 + b_1 x_h$$

$$s_{\hat{y}(new)|x_h} = \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{SSx}\right)}$$

Multiple Linear Regression

Model Statements

Population: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots. + \beta_p x_{mi} + \varepsilon_i$ $\quad$ $p = m + 1$

Sample:

$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_p x_{mi} + e_i$

$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_p x_{mi}$

$e_i = y_i - \hat{y}_i$

Confidence intervals

For β_o: $b_0 \pm t_{1-\alpha/2,n-m-1} \times s_{b_0}$

For β_j: $b_j \pm t_{1-\alpha/2,n-m-1} \times s_{b_j}$

$$\hat{y} \mid \mathbf{x}_h \pm t_{n-m-1,1-\alpha/2} \times s_{\hat{y}|\mathbf{x}_h}$$

Prediction intervals

$$\hat{y}_{(new)} \mid \mathbf{x}_h \pm t_{n-m-1,1-\alpha/2} \times s_{\hat{y}(new)|\mathbf{x}_h}$$

Test statistics (MLR only)

$$\text{partial } F = \frac{(\text{SSreg full} - \text{SSreg reduced}) / r}{\text{SSE full} / (n - m - 1) \text{ full}}$$

r = df full model – df reduced model

$F_{\text{critical}} = F_{r,\, n\text{-}m\text{-}1 \text{ full},\, 1\text{-}\alpha}$

ANOVA table

Source	df	SS	MS	F	p-value
Regression	*m*	*SSreg*	*MSreg*= *SSreg*/m	*F*= *MSreg*/*MSE*	Prob F> $F_{(m,n-m-1,1-\alpha)}$
Error	*n*-*m*-1	*SSE*	*MSE*= *SSE*/(*n* *m*-1)		
Total	*n*-1	*SSy*			

Generalized Linear Models (GLMs) – for Logistic and Poisson Regression

$y_i = g^{-1}(\mu_i) + \varepsilon_i$

$$G = \lambda_{LR} = -2(\ln L_R - lnL_U)$$

$\lambda_{LR} \sim \chi^2{}_{r,\ 1-\alpha}$

where r = number of dropped variables

Akaike's Information Criterion

$$AIC = -2\ln L + 2(k+s)$$

k = levels of y -1 s = number of predictor variables (including indicator variables)

$$Pseudo\ R^2 = 1 - \left\{\frac{L(0)}{L(\hat{\beta})}\right\}^{2/n}$$

Scaled Pseudo R^2

$$\widetilde{R^2} = \frac{R^2}{R^2\ max} \qquad R^2 \max = 1 - \{L(0)\}^{2/n}$$

Logistic Regression

Odds ratio

$$y_i = \frac{p_i}{q_i} = \frac{p_i}{1 - p_i}$$

Logit transformation

$$\ln\left(\frac{p_i}{1 - p_i}\right)$$

Linear version of the model

$$\widehat{\ln\left(\frac{p_i}{1 - p_i}\right)} = x_i{}'\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{mi}\ .$$

$$\widehat{\frac{p_i}{1 - p_i}} = \exp(x_i{}'\beta)$$

Non-linear version of the model

$$p_i = prob\widehat{(y = 1)} = \frac{\exp(x_i{}'\beta)}{1 + \exp(x_i{}'\beta)}$$

Poisson Regression

Linear version of the model

$$\ln\widehat{y_i} = x_i{}'\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{mi}$$

Non-linear version of the model

$$\widehat{y_i} = e^{x_i{}'\beta}$$

Expected probability of getting a certain count response (k)

$$P(k) = \frac{e^{-mean}\ mean^{k}}{k!}$$

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.

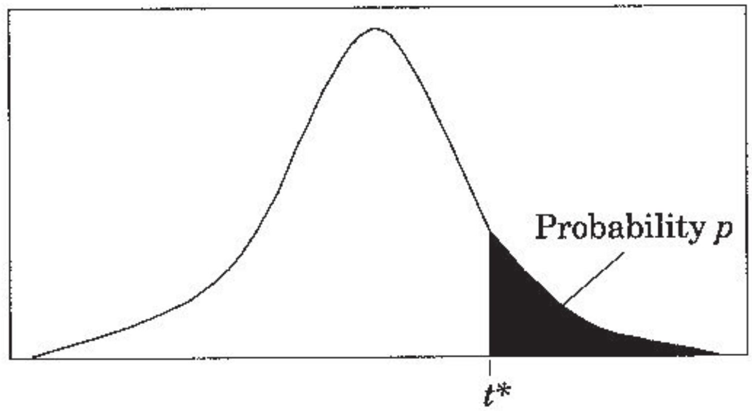

Table B t distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

F - Distribution ($\alpha = 0.05$ in the Right Tail)

Denominator Degrees of Freedom df_2 \ df_1	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2							α =	
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38