

Table of z values and probabilities for the standard normal distribution. z is the first column plus the top row. Each cell shows $P(X \leq z)$. For example $P(X \leq 1.04) = .8508$. For $z < 0$ subtract the value from 1, e.g., $P(X \leq -1.04) = 1 - .8508 = .1492$.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

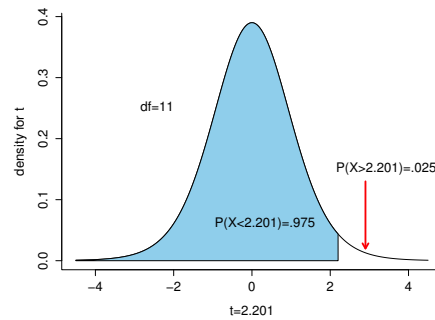


Table of t values and right tail probabilities. Degrees of freedom are in the first column (df). **Right tail probabilities** are in the first row. For example for $d.f. = 7$ and $\alpha = .05$ the critical t value for a two-tail test is 2.365 and for $d.f. = 10$ and $\alpha = .1$ the critical t value for a one-tail test is 1.372.

df	.1	.05	.025	.01	.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
50	1.299	1.676	2.009	2.403	2.678
75	1.293	1.665	1.992	2.377	2.643
100	1.290	1.660	1.984	2.364	2.626

Table of F values for right tail probabilities of .05. Degrees of freedom for denominator are in the first column and degrees of freedom for the numerator are in the top row.

denom.	numerator df									
df	1	2	3	4	5	7	10	15	50	100
1	161.45	199.5	215.71	224.58	230.16	236.77	241.88	245.95	251.77	253.04
2	18.51	19	19.16	19.25	19.3	19.35	19.4	19.43	19.48	19.49
3	10.13	9.55	9.28	9.12	9.01	8.89	8.79	8.7	8.58	8.55
4	7.71	6.94	6.59	6.39	6.26	6.09	5.96	5.86	5.7	5.66
5	6.61	5.79	5.41	5.19	5.05	4.88	4.74	4.62	4.44	4.41
6	5.99	5.14	4.76	4.53	4.39	4.21	4.06	3.94	3.75	3.71
7	5.59	4.74	4.35	4.12	3.97	3.79	3.64	3.51	3.32	3.27
8	5.32	4.46	4.07	3.84	3.69	3.5	3.35	3.22	3.02	2.97
9	5.12	4.26	3.86	3.63	3.48	3.29	3.14	3.01	2.8	2.76
10	4.96	4.1	3.71	3.48	3.33	3.14	2.98	2.85	2.64	2.59
11	4.84	3.98	3.59	3.36	3.2	3.01	2.85	2.72	2.51	2.46
12	4.75	3.89	3.49	3.26	3.11	2.91	2.75	2.62	2.4	2.35
13	4.67	3.81	3.41	3.18	3.03	2.83	2.67	2.53	2.31	2.26
14	4.6	3.74	3.34	3.11	2.96	2.76	2.6	2.46	2.24	2.19
15	4.54	3.68	3.29	3.06	2.9	2.71	2.54	2.4	2.18	2.12
16	4.49	3.63	3.24	3.01	2.85	2.66	2.49	2.35	2.12	2.07
17	4.45	3.59	3.2	2.96	2.81	2.61	2.45	2.31	2.08	2.02
18	4.41	3.55	3.16	2.93	2.77	2.58	2.41	2.27	2.04	1.98
19	4.38	3.52	3.13	2.9	2.74	2.54	2.38	2.23	2	1.94
20	4.35	3.49	3.1	2.87	2.71	2.51	2.35	2.2	1.97	1.91
21	4.32	3.47	3.07	2.84	2.68	2.49	2.32	2.18	1.94	1.88
22	4.3	3.44	3.05	2.82	2.66	2.46	2.3	2.15	1.91	1.85
23	4.28	3.42	3.03	2.8	2.64	2.44	2.27	2.13	1.88	1.82
24	4.26	3.4	3.01	2.78	2.62	2.42	2.25	2.11	1.86	1.8
25	4.24	3.39	2.99	2.76	2.6	2.4	2.24	2.09	1.84	1.78
26	4.23	3.37	2.98	2.74	2.59	2.39	2.22	2.07	1.82	1.76
27	4.21	3.35	2.96	2.73	2.57	2.37	2.2	2.06	1.81	1.74
28	4.2	3.34	2.95	2.71	2.56	2.36	2.19	2.04	1.79	1.73
29	4.18	3.33	2.93	2.7	2.55	2.35	2.18	2.03	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.33	2.16	2.01	1.76	1.7
40	4.08	3.23	2.84	2.61	2.45	2.25	2.08	1.92	1.66	1.59
60	4	3.15	2.76	2.53	2.37	2.17	1.99	1.84	1.56	1.48
100	3.94	3.09	2.7	2.46	2.31	2.1	1.93	1.77	1.48	1.39
1000	3.85	3	2.61	2.38	2.22	2.02	1.84	1.68	1.36	1.26

<p>Chapter 1: no key formulas. Chapter 2: Relative Frequency=freq. of the class / n. Approx. Class Width: =(largest value-smallest value) / number of classes. Chapter 3: sample and population means</p> $\bar{x} = \sum x_i/n \text{ and } \mu = \sum x_i/N$ <p>Weighted mean and geometric mean</p> $\bar{x} = \sum w_i x_i / w_i \text{ and } \bar{x}_g = [(x_1)(x_2) \dots (x_n)]^{1/n}.$ <p>Interquartile Range: IQR = $Q_3 - Q_1$. Population and sample variance</p> $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} \text{ and } s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$ <p>Population and sample standard deviation</p> $\sigma = \sqrt{\sigma^2} \text{ and } s = \sqrt{s^2}.$ <p>Coefficient of Variation</p> $\left(\frac{\text{Standard deviation}}{\text{Mean}} \times 100 \right) \%$ <p>z-Score: $z_i = \frac{x_i - \bar{x}}{s}$. Population and Sample Covariance</p> $\sigma_{xy} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N} \text{ and } s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$ <p>Population and Sample Pearson Correlation</p> $\rho_{xy} = \sigma_{xy} / (\sigma_x \sigma_y) \text{ and } r_{xy} = s_{xy} / (s_x s_y).$ <p>Chapter 4: Counting Rule for Combinations</p> $C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}$ <p>Counting Rule for Permutations</p> $P_n^N = n! \binom{N}{n} = \frac{N!}{(N-n)!}$ <p>Probability Rules: $P(A) = 1 - P(A^c)$</p>	<p>Chapter 4 continued:</p> $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A B) = \frac{P(A \cap B)}{P(B)}$ $P(A \cap B) = P(B)P(A B) = P(A)P(B A).$ <p>Multiplication Law for Independent Events</p> $P(A \cap B) = P(B)P(A).$ <p>Bayes' Theorem</p> $P(A_i B) = \frac{P(A_i)P(B A_i)}{P(A_1)P(B A_1) + P(A_2)P(B A_2) + \dots + P(A_n)P(B A_n)}$ <p>Chapter 5: Discrete Uniform Probability Mass Function: $f(x) = 1/n$. Expected Value of a Discrete R. V.: $E(x) = \mu = \sum x f(x)$. Variance of a Discrete R. V.:</p> $Var(x) = \sigma^2 = \sum (x - \mu)^2 f(x).$ <p>Number of Experimental Outcomes Providing Exactly x Successes in n Trials</p> $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ <p>Binomial Probability Mass Function</p> $P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}.$ <p>Expected Value for Binomial Distribution: $E(x) = \mu = np$. Variance for Binomial Distr.: $Var(x) = \sigma^2 = np(1-p)$. Poisson Probability Mass Function:</p> $P(X = x \mu) = f(x) = \frac{\mu^x e^{-\mu}}{x!}.$ <p>Hypergeometric Probability Mass Function and Expected Value:</p> $f(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \text{ and } E(x) = \mu = \frac{nr}{N}.$	<p>Chapter 5 continued: Variance for the Hypergeometric Distribution:</p> $Var(x) = \sigma^2 = n \left(\frac{r}{N} \right) \left(1 - \frac{r}{N} \right) \left(\frac{N-n}{N-1} \right).$ <p>Chapter 6: Uniform PDF</p> $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$ <p>Normal PDF The density function is</p> $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$ <p>Converting to the Standard Normal rv:</p> $z = \frac{x - \mu}{\sigma}.$ <p>Exponential PDF and CDF for $x \geq 0$</p> $f(x) = \mu^{-1} e^{-x/\mu} \text{ and } P(x \leq x_0) = 1 - e^{-x_0/\mu}.$ <p>Chapter 7: expected value of \bar{x}</p> $E(\bar{x}) = \mu.$ <p>Standard Deviation of \bar{x} (Standard Error)</p> $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$ <p>Expected Value and Std Dev (Standard Error) of \bar{p}</p> $E(\bar{p}) = p \text{ and } \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$ <p>Finite Pop. Correction Factor: $\sqrt{(N-n)/(N-1)}$. Chapter 8: Interval Estimate of Population Mean, σ known and unknown</p> $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ and } \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ <p>Necessary Sample Size for Interval Estimate of μ</p> $n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$
---	---	---

<p>Chapter 8 continued: Interval Estimate of p</p> $\hat{p} \pm z_{\alpha/2} \frac{p(1-p)}{\sqrt{n}}$ <p>Necessary Sample Size for Interval Estimate of p</p> $n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$ <p>Chapter 9: Test Statistic for Hypothesis Tests About μ, σ known and unknown</p> $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \text{ and } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ <p>Test Stat for Hypothesis About p</p> $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ <p>Chapter 10: Point Estimate and Standard Error for Difference in Two Population Means</p> $\bar{x}_1 - \bar{x}_2 \text{ and } \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ <p>Interval Estimate and Test Statistic for Difference in Two Means with Known Variances</p> $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ and } z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ <p>Interval Estimate and Test Statistic for Difference in Two Means with Unknown Variances</p> $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ and } t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ <p>Degrees of Freedom for t, Two Independent Random Samples</p> $df = \frac{1}{\frac{1}{n_1-1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1} \left(\frac{s_2^2}{n_2}\right)^2} \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$	<p>Chapter 10 continued: Test Statistic (Matched Samples)</p> $t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$ <p>ANOVA Related:</p> $\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \quad s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1} \quad \bar{\bar{x}} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}}{n_T}$ $MSTR = \frac{SSTR}{k-1} \quad SSTR = \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 \quad MSE = \frac{SSE}{n_T - k}$ $SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad F = MSTR/MSE$ $SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2 \quad SST = SSTR + SSE$ <p>Chapter 11: not covered in this course Chapter 12: $y = \beta_0 + \beta_1 x + \epsilon$</p> $E(y) = \beta_0 + \beta_1 x \quad \hat{y} = b_0 + b_1 x \quad b_0 = \bar{y} - b_1 \bar{x}$ $b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad SSE = \sum (y_i - \hat{y}_i)^2$ $SST = \sum (y_i - \bar{y})^2 \quad SSR = \sum (\hat{y}_i - \bar{y})^2 \quad SST = SSR + SSE$ $r^2 = \frac{SSR}{SST} \quad r_{xy} = (\text{sign of } b_1) \sqrt{r^2} \quad s^2 = MSE = \frac{SSE}{n-2}$ <p>Standard Error of the Estimate, $s = \sqrt{MSE}$.</p> $\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}} \quad s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} \quad t = \frac{b_1}{s_{b_1}}$ <p>For simple regression, $MSR = SSR$ because there is only one independent variable.</p> $F = \frac{MSR}{MSE} \quad s_{\hat{y}^*} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ <p>Confidence Interval for $E(y^*)$: $\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$</p> $s_{\text{pred}} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$	<p>Chapter 8 continued: Interval Estimate of p</p> $\hat{p} \pm z_{\alpha/2} \frac{p(1-p)}{\sqrt{n}}$ <p>Necessary Sample Size for Interval Estimate of p</p> $n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$ <p>Chapter 9: Test Statistic for Hypothesis Tests About μ, σ known and unknown</p> $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \text{ and } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ <p>Test Stat for Hypothesis About p</p> $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ <p>Chapter 10: Point Estimate and Standard Error for Difference in Two Population Means</p> $\bar{x}_1 - \bar{x}_2 \text{ and } \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ <p>Interval Estimate and Test Statistic for Difference in Two Means with Known Variances</p> $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ and } z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ <p>Interval Estimate and Test Statistic for Difference in Two Means with Unknown Variances</p> $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ and } t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ <p>Degrees of Freedom for t, Two Independent Random Samples</p> $df = \frac{1}{\frac{1}{n_1-1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1} \left(\frac{s_2^2}{n_2}\right)^2} \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$
<p>Chapter 10 continued: Test Statistic (Matched Samples)</p> $t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$ <p>ANOVA Related:</p> $\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \quad s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1} \quad \bar{\bar{x}} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}}{n_T}$ $MSTR = \frac{SSTR}{k-1} \quad SSTR = \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 \quad MSE = \frac{SSE}{n_T - k}$ $SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad F = MSTR/MSE$ $SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2 \quad SST = SSTR + SSE$ <p>Chapter 11: not covered in this course Chapter 12: $y = \beta_0 + \beta_1 x + \epsilon$</p> $E(y) = \beta_0 + \beta_1 x \quad \hat{y} = b_0 + b_1 x \quad b_0 = \bar{y} - b_1 \bar{x}$ $b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad SSE = \sum (y_i - \hat{y}_i)^2$ $SST = \sum (y_i - \bar{y})^2 \quad SSR = \sum (\hat{y}_i - \bar{y})^2 \quad SST = SSR + SSE$ $r^2 = \frac{SSR}{SST} \quad r_{xy} = (\text{sign of } b_1) \sqrt{r^2} \quad s^2 = MSE = \frac{SSE}{n-2}$ <p>Standard Error of the Estimate, $s = \sqrt{MSE}$.</p> $\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}} \quad s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} \quad t = \frac{b_1}{s_{b_1}}$ <p>For simple regression, $MSR = SSR$ because there is only one independent variable.</p> $F = \frac{MSR}{MSE} \quad s_{\hat{y}^*} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ <p>Confidence Interval for $E(y^*)$: $\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$</p> $s_{\text{pred}} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$	<p>Chapter 12 continued: Prediction Interval for y^*:</p> $\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$ <p>Residual for Observation i: $y_i - \hat{y}_i$</p> <p>Chapter 13:</p> $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$ $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$ $SST = SSR + SSE \quad R^2 = \frac{SSR}{SST}$ $R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$ $MSR = \frac{SSR}{p} \quad MSE = \frac{SSE}{n-p-1} \quad F = \frac{MSR}{MSE}$ $t = \frac{b_i}{s_{b_i}}$ <p>Other Math Rule Reminders:</p> $e^x = \exp(x)$ $\ln 1 = 0 \quad \ln e = 1$ $x! = (x)(x-1)(x-2) \dots (2)(1)$ $0! = 1 \quad x^0 = 1$	<p>Chapter 12 continued: Prediction Interval for y^*:</p> $\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$ <p>Residual for Observation i: $y_i - \hat{y}_i$</p> <p>Chapter 13:</p> $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$ $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$ $SST = SSR + SSE \quad R^2 = \frac{SSR}{SST}$ $R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$ $MSR = \frac{SSR}{p} \quad MSE = \frac{SSE}{n-p-1} \quad F = \frac{MSR}{MSE}$ $t = \frac{b_i}{s_{b_i}}$ <p>Other Math Rule Reminders:</p> $e^x = \exp(x)$ $\ln 1 = 0 \quad \ln e = 1$ $x! = (x)(x-1)(x-2) \dots (2)(1)$ $0! = 1 \quad x^0 = 1$

Use the output below to answer the appropriate questions in the exam. Be sure to check which table the question references.

ANOVA Table 1

	SS	DF	MS	F	p
Treatments	31.162	4	7.791	2.657	0.0712
Error	46.91	16	2.932		

Here is regression output from a model where attendance is the independent variable and regular course points (excludes extra credit from polls) is the dependent variable. The data are from my Spring 2016 classes where I am using poll everywhere points as a proxy for attendance. Many students choose not to come to class but I rather think attending the lectures is helpful. Let's see what the data say.

Regression Table 2

Regression Statistics	
R Square	0.2196685
Adjusted R Square	0.2185966
Standard Error	99.0777428
Observations	730

ANOVA

	SS	DF	MS	F	p
Regression	2011741.4	1	2011741.4	204.937	0
Residual	7146338.563	728	9816.399		

	coefficients	standard error	t stat	p-value
intercept	717.4464638	6.4083776	111.95446	0
attendance		0.4687815	14.31561	0

VERSION A

Choose the best answer. Do not write letters in the margin or communicate with other students in any way. If you have a question note it on your exam and ask for clarification when your exam is returned. In the meantime choose the best answer. Neither the proctors nor Dr. Cox will answer questions during the exam.

Please check each question and possible answers thoroughly as questions at the bottom of a page sometimes run onto the next page. Please verify that your test version and scantron version are the same.

This exam has 25 questions.

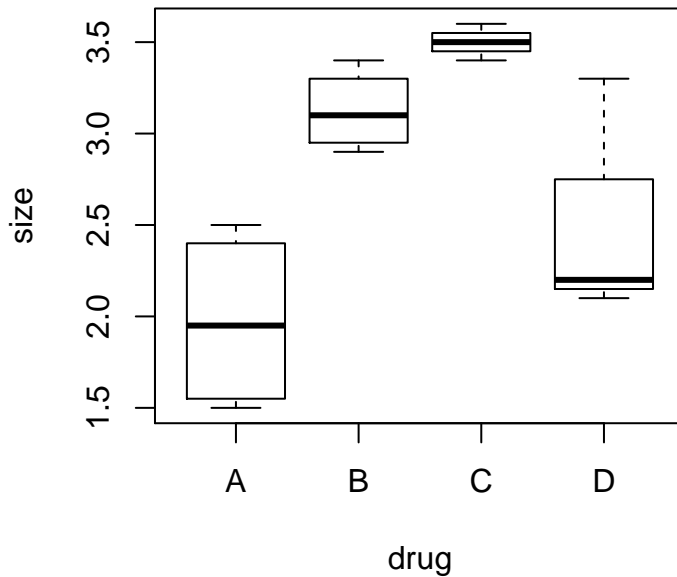
1. I have checked that my ID is bubbled in correctly. If it is bubbled in incorrectly I will get this question wrong. I also understand that questions and their possible answers may run onto the next page and so I should always check the top of the next page for possible answers. I understand that if I have a question I should simply make a note on my exam and ask Dr. Cox afterwards. I should always choose the best answer.

- (a) False.
- (b) I didn't read the directions.
- (c) True.

2. Consider the null hypothesis $H_0 : \mu_1 = \mu_2$ and a separate hypothesis $H_0 : \mu_1 = \mu_2 + .5$. Suppose that you are going to conduct hypothesis tests for both of those hypotheses at the .05 level of significance. The critical value(s) for each test is (are) the same.

- (a) true
- (b) false

3. The graphs shown here are



- (a) histograms
 - (b) scatter plots
 - (c) stem and leaf displays
 - (d) box and whiskers plots
4. The probability that a z value is less than 0.5 is
- (a) 0.5
 - (b) 0.8413447
 - (c) 0.4640237
 - (d) 0.6616608
 - (e) 0.6914625
5. Consult Table 1. From the table you can conclude that the total sum of squares is ?
- (a) 46.91
 - (b) 78.072381
 - (c) 15.747619
 - (d) 39.0361905
6. Consult Table 1. From the table you can conclude that the total number of observations used in this analysis/experiment was?
- (a) 19
 - (b) 21
 - (c) 4
 - (d) 16
7. Consult Table 1. From the table what can you conclude concerning the null hypothesis?
- (a) fail to reject
 - (b) reject the null
 - (c) cannot be determined
 - (d) depends on the number of observations.
8. Consult Table 1. What is the test statistic?

- (a) 0.071234.
 - (b) 7.7905952.
 - (c) 2.6572058.
 - (d) 1.3286029.
9. Consult Table 1. What is the critical value for a test at the .05 level?
- (a) 1.96
 - (b) 3.84
 - (c) 6.39
 - (d) 3.01
10. Suppose that the number of homework assignments a professor gives follows a Poisson distribution with a mean of 10. What is the probability of drawing a professor that gives exactly 12 assignments?
- (a) 0.012764
 - (b) 0.1143679
 - (c) 0.072765
 - (d) 0.0947803
11. Suppose that you collect data on apartment prices in Tempe. You look at 36 different apartments and find a mean of 738 and a standard deviation of 193.5. Construct a 95% confidence interval for the mean apartment price. The interval is
- (a) [672.5290193, 803.4709807]
 - (b) [676.5290193, 799.4709807]
 - (c) [666.5290193, 809.4709807]
 - (d) [605.2761174, 723.1238826]
12. Suppose that you collect data on apartment prices in Tempe. You look at 36 different apartments and find a mean of 738 and a standard deviation of 193.5. Test the hypothesis, $H_0 : \mu = 705$ at the .05 level of significance.
- (a) the test statistic is 1.0232558 so we reject the null
 - (b) the test statistic is 0.7409302 so we fail to reject
 - (c) the test statistic is 1.0232558 so we fail to reject
 - (d) the test statistic is 1.2355814 so we reject the null

13. Suppose you have a random variable that is exponentially distributed with a mean of 50. What is the probability of observing a random variable drawn from this distribution with a value of less than 50?
- (a) 0.6321206
 - (b) 0.6988058
 - (c) 0.550671
 - (d) 0.8646647
14. The greater the value of α the greater the risk of committing a Type II error.
- (a) True.
 - (b) False.
15. Consider the regression output in Table 2. What is the estimated increase in total points that a student earns (on average) when attending 1 additional lecture?
- (a) 5.1622272
 - (b) 99.0777428
 - (c) 6.7108954
 - (d) 204.937
 - (e) 9.5869935
16. Consider the regression output in Table 2. What is the estimated variance of the error term?
- (a) 9816.3991246
 - (b) 49.5388714
 - (c) 99.0777428
 - (d) 204.937
17. Consider the regression output in Table 2. What is the percentage of variation in total regular class points that can be explained by the variation in attendance?
- (a) -20.9668468%
 - (b) 26.3602162%
 - (c) 21.9668468%
 - (d) 21.8596585%

18. Consider the regression output in Table 2. What is the predicted or estimated total regular class points for a student who came 15 times?
- (a) 838.2425814
 - (b) 885.2188493
 - (c) 818.1098951
 - (d) 804.6881043
 - (e) 797.9772088
19. Consider the regression output in Table 2. Suppose you want to test the hypothesis that attendance has no impact on grades, i.e. $H_0 : \beta_1 = 0$. What is the test statistic for this hypothesis?
- (a) 14.3156138
 - (b) 17.6082049
 - (c) 0.4687815
 - (d) 0
 - (e) 111.9544623
20. Consider the regression output in Table 2. Suppose you want to test the hypothesis that attendance has no impact on grades, i.e. $H_0 : \beta_1 = 0$. What is your conclusion for this hypothesis test? (Use $\alpha = .05$.)
- (a) This is inconclusive unless we know whether it is a right tail or a left tail test.
 - (b) This cannot be determined without the appropriate df.
 - (c) reject the null
 - (d) fail to reject
21. Suppose that for all values of your explanatory variable that are above the 75th percentile you notice that the residuals are positive.
- (a) This cannot happen because the residuals must average to 0.
 - (b) The relationship between the X and Y variables is probably not linear.
 - (c) There is an indicator that there may be an issue with heteroskedasticity.
 - (d) This is an indicator that the error terms are probably correlated.
 - (e) none of the above.

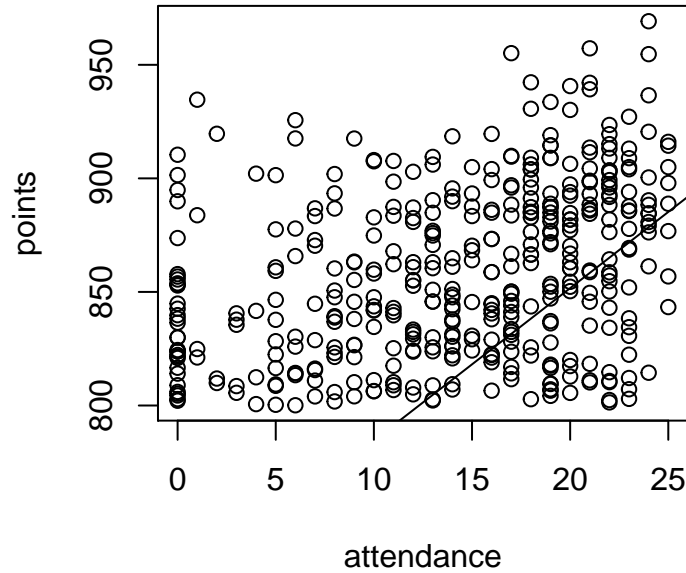
22. You are in a business meeting and a colleague presents regression results in the following form

$$\hat{y} = 12 + 2.3X.$$

You know that the regression was performed with over 100,000 observations and your colleague reports that the p-value was .011 so that it is statistically significant. However, the executive to whom you are presenting asks what the standard error was. The value is not in your power point slides but you tell her you will give it to her in just a moment. While your colleague moves to the next slide you find that it is:

- (a) 1.0618639.
 - (b) 0.4719395.
 - (c) 1.2585053.
 - (d) 0.9045507.
 - (e) 1.4158185.
23. In our treatment of regression analysis we made which assumption(s)
- (a) Y is normally distributed.
 - (b) X is normally distributed.
 - (c) ϵ is normally distributed.
 - (d) ϵ and X are normally distributed.
 - (e) all of the above.
24. The following graph shows the total points and poll everywhere points together. What kind of graph is this?

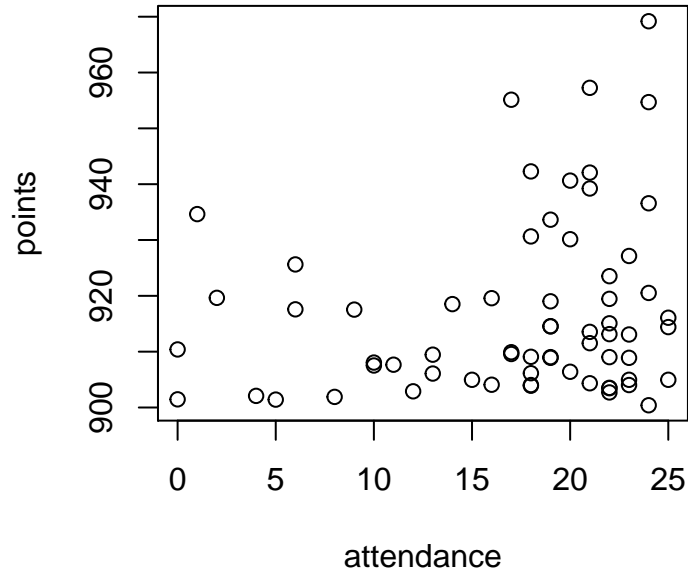
students with at least 800 points



- (a) scatter plot
- (b) bar chart
- (c) dot plot
- (d) stem and leaf display

25. The line that is superimposed in the graph **above** is the regression line from Table 2. Looking at the graph **below** concerning total class points and poll everywhere points (a proxy for attendance), which of the following is most accurate? Note: the graph below uses a subset of the observations in the graph above.

students with at least 900 points



- (a) The residuals are normally distributed
- (b) The residuals are not normally distributed.
- (c) Students with 15-20 times attending tend to do better in the class than the regression line predicts.
- (d) Students that got over 900 points (A range) attended over half the time on average.

Key

1. c
2. a
3. d
4. e
5. b
6. b
7. a
8. c
9. d
10. d
11. a
12. c
13. a
14. b, α refers to the probability of a Type I error not a Type II error.
15. c for all of the questions requiring reading the regression output you obviously need to know how to read the output in excel as shown in class. For this one you needed to work backwards from the test statistic and the standard error.
16. a,
17. c
18. c
19. a
20. c
21. b, this suggests that the relationship might be different at the 75th percentile and beyond and so it is not linear through out.

22. d, with over 100,000 observations the t distribution is essentially the same as the z distribution. If $p = .011$ the you can look on the z table and find that the test statistic was 2.54 and then you back this out to find a standard error of $2.3/2.54$ is approximately 0.9055118.
23. c
24. a
25. d