

# TABLES AND FORMULAS FOR MOORE

## *Basic Practice of Statistics*

### **Exploring Data: Distributions**

- Look for overall pattern (shape, center, spread) and deviations (outliers).

- Mean (use a calculator):

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum x_i$$

- Standard deviation (use a calculator):

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

- Median: Arrange all observations from smallest to largest. The median  $M$  is located  $(n + 1)/2$  observations from the beginning of this list.

- Quartiles: The first quartile  $Q_1$  is the median of the observations whose position in the ordered list is to the left of the location of the overall median. The third quartile  $Q_3$  is the median of the observations to the right of the location of the overall median.

- Five-number summary:

Minimum,  $Q_1$ ,  $M$ ,  $Q_3$ , Maximum

- Standardized value of  $x$ :

$$z = \frac{x - \mu}{\sigma}$$

### **Exploring Data: Relationships**

- Look for overall pattern (form, direction, strength) and deviations (outliers, influential observations).

- Correlation (use a calculator):

$$r = \frac{1}{n-1} \sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

- Least-squares regression line (use a calculator):  
 $\hat{y} = a + bx$  with slope  $b = rs_y/s_x$  and intercept  
 $a = \bar{y} - b\bar{x}$

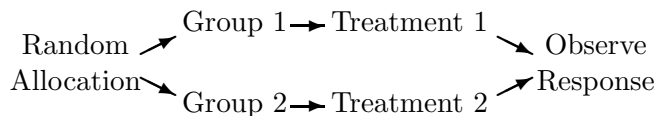
- Residuals:

$$\text{residual} = \text{observed } y - \text{predicted } y = y - \hat{y}$$

### **Producing Data**

- Simple random sample: Choose an SRS by giving every individual in the population a numerical label and using Table B of random digits to choose the sample.

- Randomized comparative experiments:



### **Probability and Sampling Distributions**

- Probability rules:

- Any probability satisfies  $0 \leq P(A) \leq 1$ .
- The sample space  $S$  has probability  $P(S) = 1$ .
- For any event  $A$ ,  $P(A \text{ does not occur}) = 1 - P(A)$
- If events  $A$  and  $B$  are disjoint,  $P(A \text{ or } B) = P(A) + P(B)$ .

- Sampling distribution of a sample mean:
  - $\bar{x}$  has mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$ .
  - $\bar{x}$  has a Normal distribution if the population distribution is Normal.
  - Central limit theorem:  $\bar{x}$  is approximately Normal when  $n$  is large.

## Basics of Inference

- $z$  confidence interval for a population mean ( $\sigma$  known, SRS from Normal population):

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \quad z^* \text{ from } N(0, 1)$$

- Sample size for desired margin of error  $m$ :

$$n = \left( \frac{z^* \sigma}{m} \right)^2$$

- $z$  test statistic for  $H_0 : \mu = \mu_0$  ( $\sigma$  known, SRS from Normal population):

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad P\text{-values from } N(0, 1)$$

## Inference About Means

- $t$  confidence interval for a population mean (SRS from Normal population):

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} \quad t^* \text{ from } t(n-1)$$

- $t$  test statistic for  $H_0 : \mu = \mu_0$  (SRS from Normal population):

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad P\text{-values from } t(n-1)$$

- Matched pairs: To compare the responses to the two treatments, apply the one-sample  $t$  procedures to the observed differences.

- Two-sample  $t$  confidence interval for  $\mu_1 - \mu_2$  (independent SRSs from Normal populations):

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

with conservative  $t^*$  from  $t$  with df the smaller of  $n_1 - 1$  and  $n_2 - 1$  (or use software).

- Two-sample  $t$  test statistic for  $H_0 : \mu_1 = \mu_2$  (independent SRSs from Normal populations):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with conservative  $P$ -values from  $t$  with df the smaller of  $n_1 - 1$  and  $n_2 - 1$  (or use software).

## Inference About Proportions

- Sampling distribution of a sample proportion: when the population and the sample size are both large and  $p$  is not close to 0 or 1,  $\hat{p}$  is approximately Normal with mean  $p$  and standard deviation  $\sqrt{p(1-p)/n}$ .

- Large-sample  $z$  confidence interval for  $p$ :

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad z^* \text{ from } N(0, 1)$$

Plus four to greatly improve accuracy: use the same formula after adding 2 successes and two failures to the data.

- $z$  test statistic for  $H_0 : p = p_0$  (large SRS):

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad P\text{-values from } N(0, 1)$$

- Sample size for desired margin of error  $m$ :

$$n = \left( \frac{z^*}{m} \right)^2 p^*(1-p^*)$$

where  $p^*$  is a guessed value for  $p$  or  $p^* = 0.5$ .

- Large-sample  $z$  confidence interval for  $p_1 - p_2$ :

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \text{SE} \quad z^* \text{ from } N(0, 1)$$

where the standard error of  $\hat{p}_1 - \hat{p}_2$  is

$$\text{SE} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Plus four to greatly improve accuracy: use the same formulas after adding one success and one failure to each sample.

- Two-sample  $z$  test statistic for  $H_0 : p_1 = p_2$  (large independent SRSs):

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where  $\hat{p}$  is the pooled proportion of successes.

## The Chi-Square Test

- Expected count for a cell in a two-way table:

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{\text{table total}}$$

- Chi-square test statistic for testing whether the row and column variables in an  $r \times c$  table are unrelated (expected cell counts not too small):

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

with  $P$ -values from the chi-square distribution with  $\text{df} = (r - 1) \times (c - 1)$ .

- Describe the relationship using percents, comparison of observed with expected counts, and terms of  $X^2$ .

## Inference for Regression

- The regression model: We have  $n$  observations on  $x$  and  $y$ . The response  $y$  for any fixed  $x$  has a Normal distribution with mean given by the true regression line  $\mu_y = \alpha + \beta x$  and standard deviation  $\sigma$ . Parameters are  $\alpha$ ,  $\beta$ ,  $\sigma$ .
- Estimate  $\alpha$  by the intercept  $a$  and  $\beta$  by the slope  $b$  of the least-squares line. Estimate  $\sigma$  by the regression standard error:

$$s = \sqrt{\frac{1}{n - 2} \sum \text{residual}^2}$$

Use software for all standard errors in regression.

- $t$  confidence interval for regression slope  $\beta$ :

$$b \pm t^* \text{SE}_b \quad t^* \text{ from } t(n - 2)$$

- $t$  test statistic for no linear relationship,  $H_0 : \beta = 0$ :

$$t = \frac{b}{\text{SE}_b} \quad P\text{-values from } t(n - 2)$$

- $t$  confidence interval for mean response  $\mu_y$  when  $x = x^*$ :

$$\hat{y} \pm t^* \text{SE}_{\hat{\mu}} \quad t^* \text{ from } t(n - 2)$$

- $t$  prediction interval for an individual observation  $y$  when  $x = x^*$ :

$$\hat{y} \pm t^* \text{SE}_{\hat{y}} \quad t^* \text{ from } t(n - 2)$$

## One-way Analysis of Variance: Comparing Several Means

- ANOVA  $F$  tests whether all of  $I$  populations have the same mean, based on independent SRSs from  $I$  Normal populations with the same  $\sigma$ .  $P$ -values come from the  $F$  distribution with  $I - 1$  and  $N - I$  degrees of freedom, where  $N$  is the total observations in all samples.
- Describe the data using the  $I$  sample means and standard deviations and side-by-side graphs of the samples.
- The ANOVA  $F$  test statistic (use software) is  $F = \text{MSG}/\text{MSE}$ , where

$$\text{MSG} = \frac{n_1(\bar{x}_1 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2}{I - 1}$$

$$\text{MSE} = \frac{(n_1 - 1)s_1^2 + \cdots + (n_I - 1)s_I^2}{N - I}$$



TABLE B Random digits								
Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915
115	61041	77684	94322	24709	73698	14526	31893	32592
116	14459	26056	31424	80371	65103	62253	50490	61181
117	38167	98532	62183	70632	23417	26185	41448	75532
118	73190	32533	04470	29669	84407	90785	65956	86382
119	95857	07118	87664	92099	58806	66979	98624	84826
120	35476	55972	39421	65850	04266	35435	43742	11937
121	71487	09984	29077	14863	61683	47052	62224	51025
122	13873	81598	95052	90908	73592	75186	87136	95761
123	54580	81507	27102	56027	55892	33063	41842	81868
124	71035	09001	43367	49497	72719	96758	27611	91596
125	96746	12149	37823	71868	18442	35119	62103	39244

TABLE C <i>t</i> distribution critical values												
df	Upper tail probability <i>p</i>											
	.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	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
<i>z</i> *	0.674	0.841	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 <i>C</i>											

df	Upper tail probability $p$											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

$n$	Upper tail probability $p$									
	.20	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
3	0.8090	0.9511	0.9877	0.9969	0.9980	0.9995	0.9999	1.0000	1.0000	1.0000
4	0.6000	0.8000	0.9000	0.9500	0.9600	0.9800	0.9900	0.9950	0.9980	0.9990
5	0.4919	0.6870	0.8054	0.8783	0.8953	0.9343	0.9587	0.9740	0.9859	0.9911
6	0.4257	0.6084	0.7293	0.8114	0.8319	0.8822	0.9172	0.9417	0.9633	0.9741
7	0.3803	0.5509	0.6694	0.7545	0.7766	0.8329	0.8745	0.9056	0.9350	0.9509
8	0.3468	0.5067	0.6215	0.7067	0.7295	0.7887	0.8343	0.8697	0.9049	0.9249
9	0.3208	0.4716	0.5822	0.6664	0.6892	0.7498	0.7977	0.8359	0.8751	0.8983
10	0.2998	0.4428	0.5494	0.6319	0.6546	0.7155	0.7646	0.8046	0.8467	0.8721
11	0.2825	0.4187	0.5214	0.6021	0.6244	0.6851	0.7348	0.7759	0.8199	0.8470
12	0.2678	0.3981	0.4973	0.5760	0.5980	0.6581	0.7079	0.7496	0.7950	0.8233
13	0.2552	0.3802	0.4762	0.5529	0.5745	0.6339	0.6835	0.7255	0.7717	0.8010
14	0.2443	0.3646	0.4575	0.5324	0.5536	0.6120	0.6614	0.7034	0.7501	0.7800
15	0.2346	0.3507	0.4409	0.5140	0.5347	0.5923	0.6411	0.6831	0.7301	0.7604
16	0.2260	0.3383	0.4259	0.4973	0.5177	0.5742	0.6226	0.6643	0.7114	0.7419
17	0.2183	0.3271	0.4124	0.4821	0.5021	0.5577	0.6055	0.6470	0.6940	0.7247
18	0.2113	0.3170	0.4000	0.4683	0.4878	0.5425	0.5897	0.6308	0.6777	0.7084
19	0.2049	0.3077	0.3887	0.4555	0.4747	0.5285	0.5751	0.6158	0.6624	0.6932
20	0.1991	0.2992	0.3783	0.4438	0.4626	0.5155	0.5614	0.6018	0.6481	0.6788
30	0.1594	0.2407	0.3061	0.3610	0.3770	0.4226	0.4629	0.4990	0.5415	0.5703
40	0.1368	0.2070	0.2638	0.3120	0.3261	0.3665	0.4026	0.4353	0.4741	0.5007
50	0.1217	0.1843	0.2353	0.2787	0.2915	0.3281	0.3610	0.3909	0.4267	0.4514
60	0.1106	0.1678	0.2144	0.2542	0.2659	0.2997	0.3301	0.3578	0.3912	0.4143
80	0.0954	0.1448	0.1852	0.2199	0.2301	0.2597	0.2864	0.3109	0.3405	0.3611
100	0.0851	0.1292	0.1654	0.1966	0.2058	0.2324	0.2565	0.2786	0.3054	0.3242
1000	0.0266	0.0406	0.0520	0.0620	0.0650	0.0736	0.0814	0.0887	0.0976	0.1039