

---

Student's Name:

SFU id number:

---

Please write your name on every page. Please do not open the exam until you are instructed to do so.

No aids are permitted except for standard scientific calculators without Wi- Fi, Bluetooth or cellular capabilities.

No cell phones, PDAs, media players, or other electronic items are permitted.

Notes are not permitted

Normal tables appear at the end of the exam package. Feel free to remove them

page	value
2	8
5	8
6	13
7	7 + 2 bonus
8	10
11	15
12	8
15	10
16	7
17	4 + 2 bonus
18	9

total: 99 + 4 bonus.

Each of us has an ABO blood type which describes whether two characteristics called A and B are present. Every human being has 2 blood type alleles (gene forms), one inherited from the mother and one from the father. Each of these alleles can be A, B or O. Which two we inherit determines our blood type. The table below shows what our blood type is for each combination of two alleles. We inherit each of a parents two alleles with probability 0.5. We inherit independently from our mother and father.

Alleles inherited	Blood type
A and A	A
A and B	AB
A and O	A
B and B	B
B and O	B
O and O	O

A mother has alleles A and O and a father has alleles A and B.

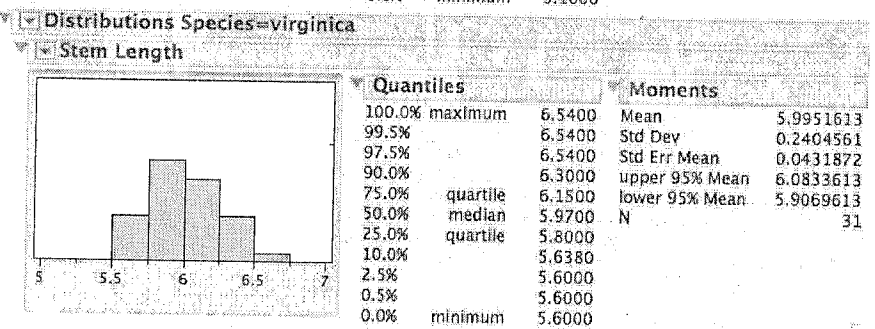
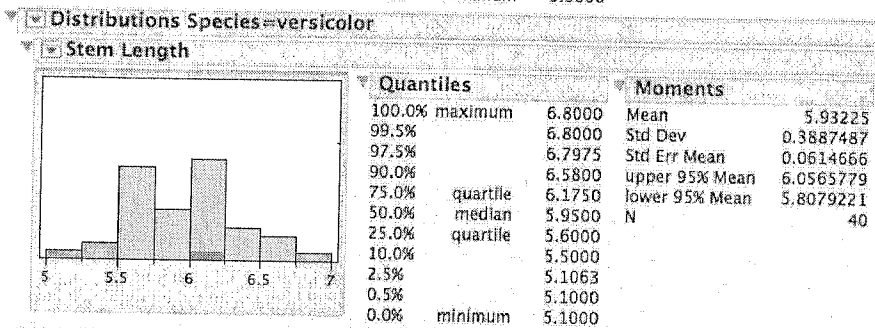
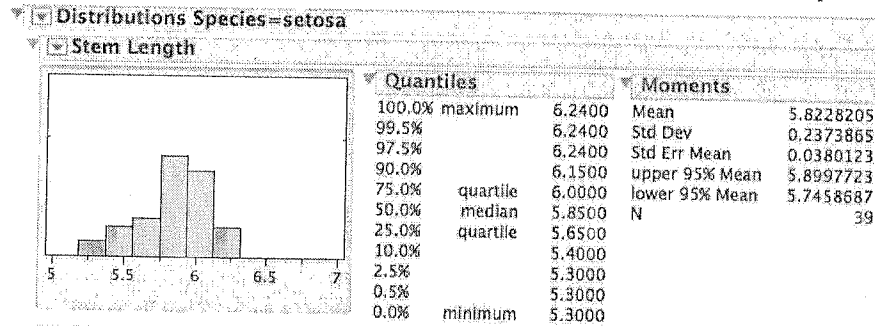
a) What are the blood types that their children could potentially have? (1pt)

b) What is the probability that their child has blood type A? (2pt)

c) What is the probability that their child has blood type AB? (2pt)

d) If they have two children what is the probability that both children have the blood type A? (3pt)

A researcher wants to know if the mean iris stem length varies by species. The JMP output is included below. Use this information for the questions on the next pages. Feel free to remove this page if it helps. This output is necessary for pages 5-7.



**Oneway Anova**

**Summary of Fit**

Rsquare	0.052458
Adj Rsquare	0.034747
Root Mean Square Error	0.302171
Mean of Response	5.911182
Observations (or Sum Wgts)	110

**Analysis of Variance**

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Species		0.540885	0.270442	2.9619	0.0560
Error		9.769861	0.091307		
C. Total		10.310746			

**Means for Oneway Anova**

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
setosa	39	5.82282	0.04839	5.7269	5.9187
versicolor	40	5.93225	0.04778	5.8375	6.0270
virginica	31	5.99516	0.05427	5.8876	6.1027

Std Error uses a pooled estimate of error variance

Scrap page

---

Using the iris dataset,

- a) draw a boxplot for the setosa species Stem Length. Include any outliers if they exist and also mark the location of the mean. Since the original data is not given, if you find outliers, draw only those whose values that you can actually determine from the JMP output. ( 5 pts for the basic boxplot + 1 pt for marking the mean + 2 pts for justifying your conclusion about outliers and if applicable including them in the picture.)

---

Using the output for the iris data

b) (3pts) From the OneWay ANOVA table JMP output, what are the missing values of the df for:

Species \_\_\_\_\_ Error \_\_\_\_\_ C. Total \_\_\_\_\_

c) A researcher wants to know if the mean iris stem length varies by species. Use the 4 step plan to provide an answer for the researcher (1pt for state, 3 pts for plan, 3 pts for solve, 3pts for conclude, total 10 pts)

---

Using the output for the iris data

- b) Look at the JMP output for the Versicolor species from the iris data set. Using a normal approximation to the distribution of individuals, what does that Normal model say the inter-quartile range should be? ( 4pts)

- c) Look at the JMP output for the Virginica species from the iris data set. Using a normal approximation to the distribution, what is the probability that a stem length is less than 5.76? ( 3pts)

Bonus (2pts) In which Canadian province was the original Iris data set (seen in class) collected?

---

This page has 3 multiple choice questions. Circle the correct response. Multiple choice questions are worth 2 pts each

A study attempts to determine whether a football filled with helium when kicked travels farther than one that is filled with air. Each subject kicks twice—once with a football filled with helium and once with a football filled with air. The order of the type of football kicked is randomized. This is an example of

- a. a matched pairs experiment.
- b. a double-blind observational study.
- c. a stratified analysis.
- d. the placebo effect.

These next questions are all based on this one study:

One hundred volunteers who suffer from agoraphobia are available for a study. Fifty are selected at random and are given the drug imipramine, which is believed to be effective in treating agoraphobia. The other 50 are given a placebo. A psychiatrist evaluates the symptoms of all volunteers after two months to determine if there has been substantial improvement in the severity of the symptoms.

This study would be double-blind if

- a. neither drug had any identifying marks on it.
- b. neither the volunteers nor the psychiatrist were allowed to see each other during the session during which the psychiatrist evaluated the severity of the symptoms.
- c. neither the volunteers nor the psychiatrist knew which subjects had received the placebo.
- d. all of the above.

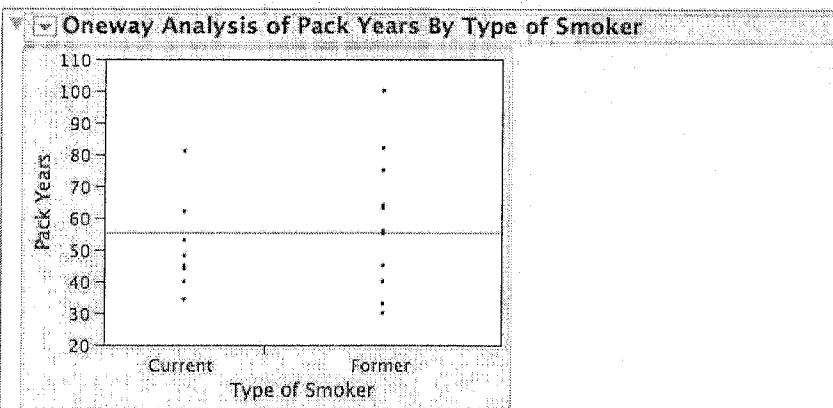
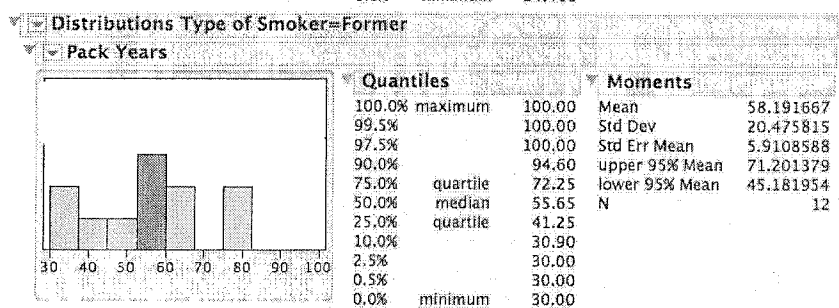
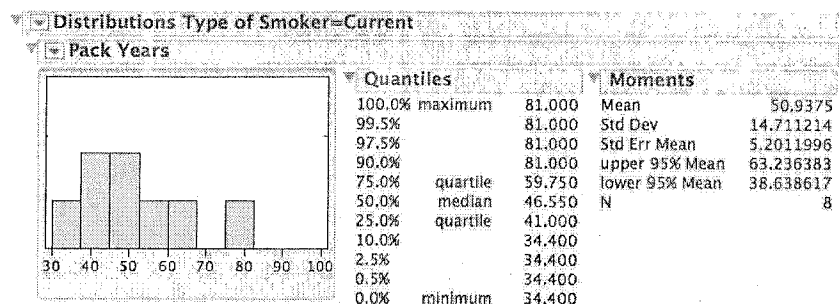
Suppose the volunteers were first divided into men and women, and then half of the men were randomly assigned to the new drug and half of the women were assigned to the new drug. The remaining volunteers received the placebo. This would be an example of

- a. replication.
- b. confounding. The effects of gender will be mixed up with the effects of the drugs.
- c. a block design.
- d. a matched pairs design.

Describe how the above agoraphobia study could be made into a cross-over design (4pts.)



Researchers randomly selected a group of 8 current and 12 former smokers from the local medical records. The individuals were asked about the total pack years in which they smoked. Pack years is a measure of the amount and the duration over which an individual smoked. The researchers suspect that people with higher pack years are more likely to quit, and would like to use their data to test if this might be true on average. The JMP output is below and the next pages use this information. Feel free to remove this page. Questions relating to this data are on page 11

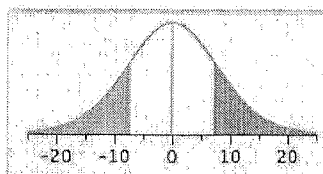


**t Test**

Former-Current

Assuming unequal variances

Difference	7.254	t Ratio	0.921349
Std Err Dif	7.873	DF	
Upper CL Dif	23.807	Prob >  t	
Lower CL Dif	-9.299	Prob > t	
Confidence	0.95	Prob < t	



---

Scrap page

---

Using the smoking data set

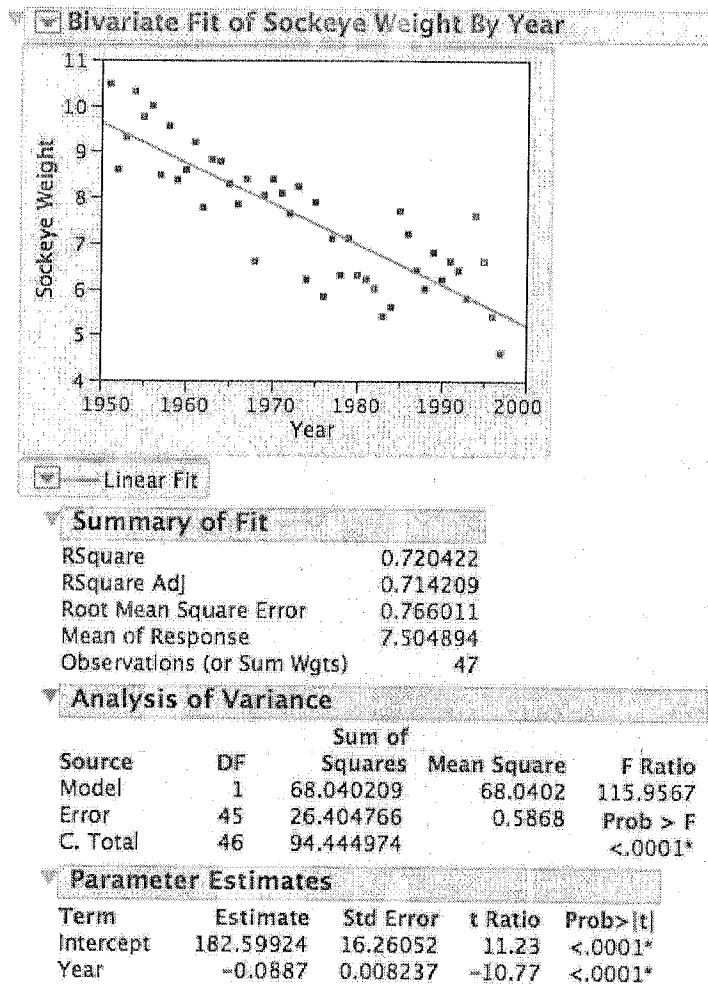
- a) Is this an experiment or observational study? ( 1pt)
- b) Is this a matched pair or two sample study? (1pt)
- c) State the hypotheses that the researcher would like analyzed?(2pts)
- d) How many degrees of freedom would you use? ( 1pt)
- e) What would you conclude? (2 pts for giving your answer in a sentence relating to the original research problem + 3pts for the p-value)
- f) Assume for the moment that you were wrong about the number of degrees of freedom in part d. Is it better to suggest too small a value for df or too large a value for df? Why? (1 pt for the right choice. 4 pts for an explanation about the impact of your choice in the context of the hypotheses in part c)

---

The resting heart rate of a Guinea Pig is 170 beats per minute with a 95% confidence interval of 160 to 190. A researcher would like to know if the actual mean heart rate could be 200 beats per minute. What would you tell that researcher? ( 1 pt for the answer 3 pts for a good explanation as to how you know.)

Give a definition for the p-value (4pts)

The weights of Sockeye salmon in a fishing region of BC are plotted against year. The regression output is below. The next questions relate to this output. Feel free to remove this page. Questions regarding this data set are on pages 15-17



---

scrap page

### Using the Sockeye data

- a) What is the regression model? (Use actual numbers and variable names) ( 3pts)
- b) Which variable is the response? ( 1pt)
- c) What proportion of the variation in sockeye salmon weight can be explained by the least squares line? ( 1pt)
- d) What is the value of the correlation in this data set? ( 2pts)
- e) Is there significant evidence that the correlation is non-zero? ( 2pts for conclusion, 1 pt for p-value)

---

Using the Sockeye data

f) According to our model, how large were the sockeye back in year 0? ( 1pt for the value)

g) How do you interpret the value of the parameter estimate for 'year' in the JMP output? ( 2pts)

h) According to our model, what is your best guess at the true size of sockeye salmon in the year 2050? (1pt) Is this value reasonable? why or why not? (3pts for explanation)



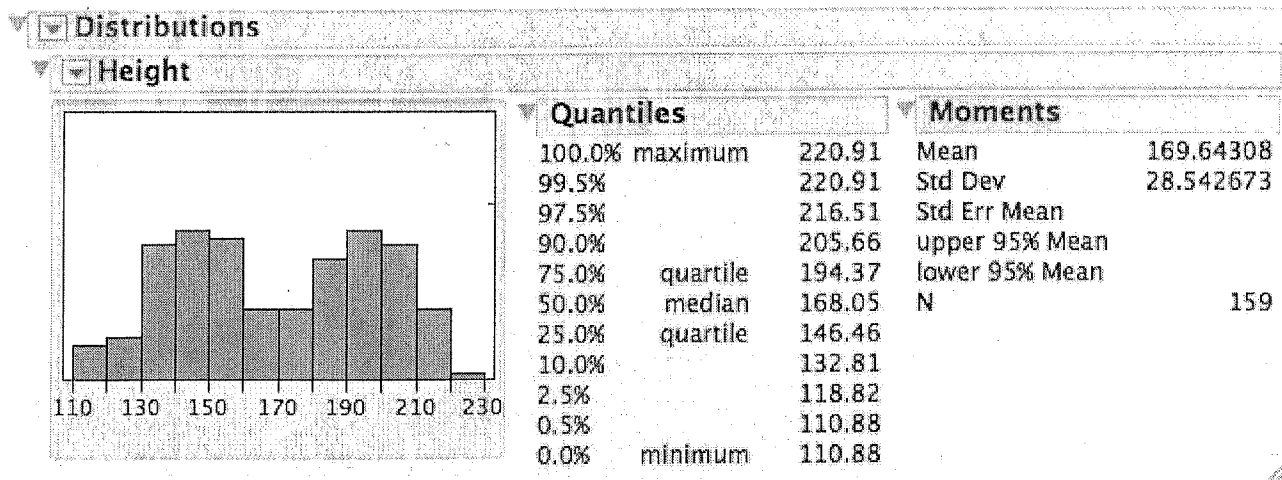
---

Using the Sockeye data

i) Draw a picture of how this data might look if it had a correlation of  $-0.9$  ( 2pts)

j) Draw a picture of how data might look if it had a correlation of 0. ( 2pts + bonus 2pts if you draw a situation where there is a relationship between variables but a correlation of 0)

The heights of a sample of students has the following histogram and JMP output. Use this output for the questions on this page.



a) Describe the shape of the distribution. ( 2pts)

b) We wish to test hypothesis that the true population mean height is actually 175 against the two sided alternative. What is the probability model that we use? ( 3 points for the name and values that define that distribution.)

c) Why can we use the probability model in part b? ( 4pts for an explanation)

Table entry for  $z$  is the area under the standard Normal curve to the left of  $z$ .

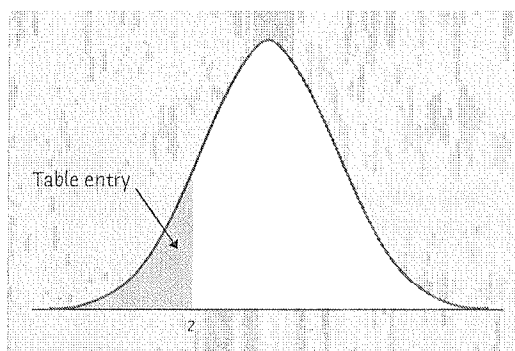


TABLE A Standard Normal probabilities

$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

A normal distribution curve is shown. The horizontal axis is marked with a point labeled '2'. The area under the curve to the left of this point is shaded. An arrow points from the text 'Table entry' to the shaded area.

[illegible]

Table entry for  $C$  is the critical value  $t^*$  required for confidence level  $C$ . To approximate one- and two-sided  $P$ -values, compare the value of the  $t$  statistic with the critical values of  $t^*$  that match the  $P$ -values given at the bottom of the table.

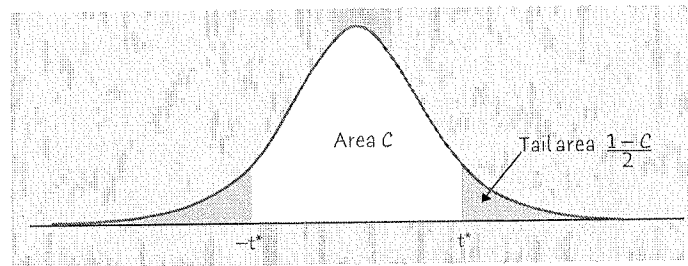


TABLE C  $t$  distribution critical values

DEGREES OF FREEDOM	CONFIDENCE LEVEL $C$											
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
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
$z^*$	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
One-sided $P$	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
Two-sided $P$	.50	.40	.30	.20	.10	.05	.04	.02	.01	.005	.002	.001

