Practice Exam 3 (Uploaded on Apr. 19, 2017) **Name:**

1. Two Web colors are used for a site advertisement. If a site visitor arrives from an affiliate, the probabilities of the blue or green colors being used in the advertisement are 0.8 and 0.2, respectively. If the site visitor arrives from a search site, the probabilities of blue and green colors in the advertisement are 0.4 and 0.6, respectively. The proportions of visitors from affiliates and search sites are 0.3 and 0.7, respectively. What is the probability that a visitor is from a search site given that the blue ad was viewed?

Sol. We define events:

$$\begin{aligned} A &= \{\text{a site visitor arrives from an affiliate}\} \\ Y &= \{\text{a site visitor arrives from a search site}\} \\ B &= \{\text{the blue colors being used in the advertisement}\} \\ G &= \{\text{the green colors being used in the advertisement}\} \end{aligned}$$

According to the statement, we have $P(B|A) = 0.8$, $P(G|A) = 0.2$, $P(B|Y) = 0.4$, and $P(G|Y) = 0.6$. Further, since the proportions of visitors from affiliates and search sites are 0.3 and 0.7, respectively, we have $P(A) = 0.3$ and $P(Y) = 0.7$. The goal is $P(Y|B)$. Since $P(Y|B) = P(Y \cap B)/P(B)$, we need to know $P(Y \cap B)$ and $P(B)$.

- For $P(Y \cap B)$, we have $P(Y \cap B) = P(B \cap Y) = P(B|Y)P(Y) = 0.4(0.7) = 0.28$.
- For $P(B)$, we use total probability rule and have $P(B) = P(B|A)P(A) + P(B|Y)P(Y) = 0.8(0.3) + 0.4(0.7) = 0.52$.

Then we conclude that $P(Y|B) = P(Y \cap B)/P(B) = 0.28/0.52 = 7/13$.

2. The phone lines to an airline reservation system are occupied 40% of the time. Assume that the events that the lines are occupied on successive calls are independent. Assume that 10 calls are placed to the airline.

(a) What is the probability that for exactly three calls, the lines are occupied?

(b) What is the probability that for at least one call, the lines are not occupied?

(c) What is the probability that for three or four calls, the lines are not occupied?

(d) What is the expected number of calls in which the lines are all occupied?

Sol.

(a)

$$X = \text{the number of lines are occupied out of 10 calls.}$$

Then $X \sim \text{Binomial}(p = 0.4, n = 10)$.

$P(X = 3) =$ `binompdf(10,0.4,3)` $= 0.21499$.

(b)

$$Y = \text{the number of lines are not occupied out of 10 calls.}$$

Then $Y \sim \text{Binomial}(p = 0.6, n = 10)$.

$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1-$ `binompdf(10,0.6,0)` $= 1 - 1.0486 \times 10^{-4}$.

(c) $P(Y = 3 \text{ or } Y = 4) = P(Y = 3) + P(Y = 4) =$ `binompdf(10,0.6,3)` + `binompdf(10,0.6,4)` $= 0.0425 + 0.1115 = 0.1540$.

Or, $P(Y = 3 \text{ or } Y = 4) = P(Y \leq 4) - P(Y \leq 2) =$ `binomcdf(10,0.6,4)` - `binomcdf(10,0.6,2)` $= 0.1662$ - $0.01229 = 0.1539$.

(d) $E(X) = np = 10(0.4) = 4$.

3. An article in *Atmospheric Chemistry and Physics* “Relationship Between Particulate Matter and Childhood Asthma—Basis of a Future Warning System for Central Phoenix” (2012, Vol 12, pp. 2479-2490) reported the use of PM10 (particulate matter $< 10\ \mu m$ diameter) air quality data measured hourly from sensors in Phoenix, Arizona. The 24-hr (daily) mean PM10 for a centrally located sensor was 50.9 $\mu g/m^3$ with a standard deviation of 25.0. Assume that the daily mean of PM10 is normally distributed.

(a) What is the probability of a daily mean of PM10 greater than $100\mu g/m^3$?

(b) What is the probability of a daily mean of PM10 less than $25\mu g/m^3$?

(c) What is the probability of a daily mean of PM10 between $25\mu g/m^3$ and $75\mu g/m^3$?

(d) What daily mean of PM10 value is exceeded with probability 5%?

Sol.

$$X = \text{daily mean of PM10.}$$

Then $X \sim$Normal$(\mu = 50.9, \sigma^2 = (25)^2)$.

(a) $P(X > 100) =$ `normalcdf(`$100, 10^{99}, 50.9, 25$`)` $= 0.02476$.

(b) $P(X < 25) =$ `normalcdf(`$-10^{99}, 25, 50.9, 25$`)` $= 0.1501$.

(c) $P(25 < X < 75) =$ `normalcdf(`$25, 75, 50.9, 25$`)` $= 0.6824$.

(d) Find x such that $0.05 = P(X \leq x)$. $x =$ `InvNorm(0.05, 50.9, 25)` $= 9.7787$.

4. Consider the hypothesis test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$. Suppose that sample sizes are $n_1 = 15$ and $n_2 = 15$, that $\overline{x}_1 = 4.7$ and $\overline{x}_2 = 7.8$, and that $s_1^2 = 4$ and $s_2^2 = 6.25$. Assume that the data are drawn from normal distributions. Use $\alpha = 0.05$.

(a) Assume that $\sigma_1 = \sigma_2$, test the hypothesis and find the P-value. What is your conclusion?

(b) Assume that $\sigma_1 \neq \sigma_2$, test the hypothesis and find the P-value. What is your conclusion?

Sol. Set significant level $\alpha = 0.05$.

(a) P-value Approach: Here P-value can be obtaining by `2-SampTTest` with `Pooled` option: **Yes**. (since population variances are unknown but equal) in calculator. Since P-value $= 8.1775 \times 10^{-4} < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the population difference, $\mu_1 - \mu_2$ is different from 0, i.e., μ_1 is different from μ_2.

(b) P-value Approach: Here P-value can be obtaining by `2-SampTTest` with `Pooled` option: **No**. (since population variances are unknown but equal) in calculator. Since P-value $= 8.6500 \times 10^{-4} < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the population difference, $\mu_1 - \mu_2$ is different from 0, i.e., μ_1 is different from μ_2.

5. An article in the the Materials Research Bulletin [1991, Vol. 26(11)] investigate four different methods of preparing the superconducting compound $PbMo_6S_8$. The authors contend that the presensce of oxygen during the preparation process affects the material's superconducting transition temperature T_c (in °K) were made for each method, and the results are as follows:

Team sheet	Transition Temperature T_c(°K)				
1	14.8	14.8	14.7	14.8	14.9
2	14.6	15.0	14.9	14.8	14.7
3	12.7	11.6	12.4	12.7	12.1
4	14.2	14.4	14.4	12.2	11.7

(a) Is there evidence to support the claim that the presence of oxygen during preparation affects the mean transition temperature? Use $\alpha = 0.05$.

(b) What is the P-value for the F-test in part (a)?

(c) According to the box plots in Figure 1 on next page, what is your intuition about the population means of transition temperature? Does that match the result in part (a)?

(d) According to Figure 1 on next page, what are your concerns about the assumption of one-way classification model? Is the assumption of equal variances reasonable? In the light of your answer here, how do you feel confortable of our decision that we made in part (a)?

Sol.

(a) Null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

versus

$$H_a : \text{Not all population means are equal.}$$

Set significant level $\alpha = 0.05$.

- P-value Appraoch: Here P-value can be obtaining by `ANOVA` in calculator. It can also be obtain by the ANOVA table provided by `R`. Since P-value$= 6.952 \times 10^{-5} < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the population means of transition temperature are different due to four different methods.

(b) P-value $= 6.952 \times 10^{-5}$.

(c) According to the box plot, since the sample means are very different, I think the population means should be different as well. My intuition matches the result in part (a).

(d) The sample variance look very different. The equal population variances assumption may not be appropriate. So I do not feel comfortable of the decision that I made in part (a).

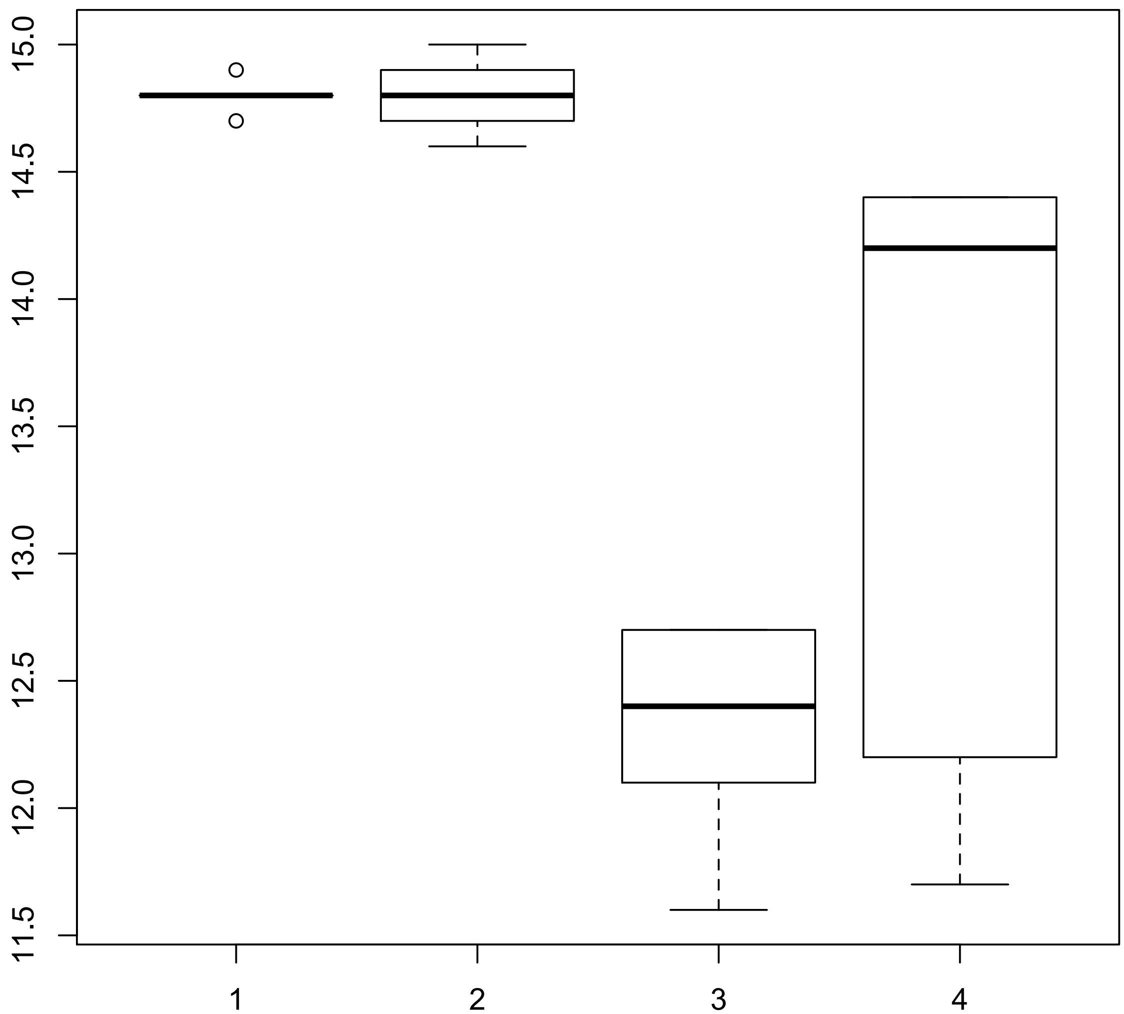

Figure 1: Box plot for superconducting compound data.

6. The journal *Human Factors* (1962, pp. 375-380) reported a study in which $n = 14$ subjects were asked to parallel park two cars having very different wheel bases and turning radii. The time in seconds for each subject was recorded and is given.

Subject	Auto1	Auto2	Difference
1	37.0	17.8	19.2
2	25.8	20.2	5.6
3	16.2	16.8	-0.6
4	24.2	41.4	-17.2
5	22.0	21.4	0.6
6	33.4	38.4	-5.0
7	23.8	16.8	7.0
8	58.2	32.2	26.0
9	33.6	27.8	5.8
10	24.4	23.2	1.2
11	23.4	29.6	-6.2
12	21.2	20.6	0.6
13	36.2	32.2	4.0
14	29.8	53.8	-24.0

(a) Please construct a 95% confidence interval for mean difference $\mu_D = \mu_1 - \mu_2$.

(b) Testing $H_0 : \mu_0 = 0$ versus $H_a : \mu_0 \neq 0$. Use significant level $\alpha = 0.05$, what is your conclusion?

Sol. According to those differences, we construct one sample confidence interval for μ_D with unknown population variance and construct one sample confidence interval `T-Interval` and perform one sample `T-Test` in calculator.

(a) We use T-Interval in calculator or confidence interval stated on page 122 in lecture to obtain the confidence interval.

Interpretation: We are 95% confident that the population mean difference μ_D of distances until tires wearing out is in $(-6.11, 8.53)$.

(b) Null and alternative hypotheses:

$$H_0 : \mu_D = 0 \text{ versus } H_a : \mu_D \neq 0.$$

Set significant level $\alpha = 0.05$.

- Confidence Approach: According to the reject criterion on page 123 in lecture notes, since 0 is in (-6.11,8.53), we do not reject the null hypothesis.
- P-value Approach: Here P-value can be obtaining by `T-Test` in calculator. Since P-value$=$ $0.7260 > 0.05 = \alpha$, we do not reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do not provide sufficient evidence to conclude that the population mean difference μ_D of parking times are 0.

7. To investigate the relationship between hydrocarbon and oxygen purity, the following data are collected.

Observation Number	Hydrocarbon Level $x(\%)$	Purity $y(\%)$	Observation Number	Hydrocarbon Level $x(\%)$	Purity $y(\%)$
1	0.99	90.01	11	1.19	93.54
2	1.02	89.05	12	1.15	92.52
3	1.15	91.43	13	0.98	90.56
4	1.29	93.74	14	1.01	89.54
5	1.46	96.73	15	1.11	89.85
6	1.36	94.45	16	1.20	90.39
7	0.87	87.59	17	1.26	93.25
8	1.23	91.77	18	1.32	93.41
9	1.55	99.42	19	1.43	94.98
10	1.40	93.65	20	0.95	87.33

The purity is the response Y, and hydrocarbon level is the regression x. A simple linear regression model is assumed as

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$. We use `R` to perform linear regression analysis and the following are the R output:

```
Call:
lm(formula = Purity ~ Hydro)

Residuals:
     Min       1Q   Median       3Q      Max
-1.83029 -0.73334  0.04497  0.69969  1.96809

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   74.283      1.593   46.62  < 2e-16 ***
Hydro         14.947      1.317   11.35 1.23e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.087 on 18 degrees of freedom
Multiple R-squared:  0.8774,Adjusted R-squared:  0.8706
F-statistic: 128.9 on 1 and 18 DF,  p-value: 1.227e-09
```

Then answer the following questions:

(a) What are $\hat{\beta}_0$, $\hat{\beta}_1$,

(b) Write down the fitted least squares regression model $\hat{Y} =$

(c) Find the estimate of σ.

(d) Find the 95% confidence interval of β_1 and give an interpretation.

(e) Testing $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. What is your conclusion?

(f) What are coefficient determination R^2 and adjust coefficient determination R^2_{adj}?

(g) According to the `R` output below, please find the 95% confidence and prediction intervals at $x = 1\%$ and give interpretations of them. Which one is larger? Why?

```
> predict(fit,data.frame(Hydro=1),level=0.95,interval='confidence')
       fit      lwr      upr
1 89.23079 88.48612 89.97547
> predict(fit,data.frame(Hydro=1),level=0.95,interval='prediction')
       fit      lwr     upr
1 89.23079 86.82969 91.6319
```

Sol.

(a) $\hat{\beta}_0 = 74.283$, $\hat{\beta}_1 = 14.947$.

(b) $\widehat{\texttt{Purity}} = 74.283 + 14.947\ \texttt{Hydro}$.

(c) $\widehat{\sigma} = 1.087$. (`Residual standard error`)

(d) $\hat{\beta}_1 \pm t_{n-2,\alpha/2} \times se(\hat{\beta}_1) = 14.947 \pm t_{n-2,\alpha/2}(1.317) = 14.947 \pm 2.100922(1.317) = (12.180, 17.714)$.
Interpretation: We are 95% confidence that the slope β_1 is between 12.180 and 17.714.

(e) Since we have P-value$= 1.23 \times 10^{-9} < 0.05$, we reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, data do provide sufficient evidence that the slope β_1 is not zero.

(f) The followings are `R` outputs:

```
> predict(fit,data.frame(Hydro=1),level=0.95,interval="confidence")
       fit      lwr      upr
1 89.23079 88.48612 89.97547
> predict(fit,data.frame(Hydro=1),level=0.95,interval="prediction")
       fit      lwr     upr
1 89.23079 86.82969 91.6319
```

When hydrocarbon level $x = 1\%$, we are 95% confident that the mean oxygen purity percentage is between 88.486 and 89.975.

When hydrocarbon level $x = 1\%$, we are 95% confident that the oxygen purity percentage for a single run of the exp will be between 86.830 and 91.632.

t Table upper-tail probability:

df	.25	.10	.05	.025	.01	.005
1	1.0000	3.0777	6.314	12.706	31.821	63.657
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.925
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800
50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778

t Table upper-tail probability:

df	.25	.10	.05	.025	.01	.005
51	0.6793	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6459
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.2931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9853	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2902	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
110	0.6767	1.2893	1.6588	1.9818	2.3607	2.6213
120	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174
∞	0.6745	1.2816	1.6449	1.9600	2.3264	2.5758