1. (4 pts.) Let $X_1, ..., X_n$ and $Y_1, ..., Y_n$ denote two independent normal random samples from the distributions $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$ where σ^2 is known. Find n such that

$$P(\bar{X} - \bar{Y} - \sigma/5 < \mu_X - \mu_Y < \bar{X} - \bar{Y} + \sigma/5) = 0.90.$$

We have that $1.645\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}} = \sigma/5$ which gives $n = 136$ (rounded up).

2. (6 pts.) Consider the following data on job satisfaction and income. Test the null hypothesis that the distribution of incomes is the same between satisfied and dissatisfied workers at level $\alpha = 0.10$.

	Job Satisfaction	
Income	Satisfied	Dissatisfied
Low	20	25
Middle	23	17
High	31	9

The expected cell counts (from left to right and top to bottom) are given by $74(45/125) = 26.64$, 18.36, 23.68, 16.32, 23.68, 16.32. The chi square test statistic has value 9.65 and 2 degrees of freedom. This is greater than 4.605 so reject H_0. The p-value is less than 0.01.

3. (8 pts.) For $i = 1, ..., n$ let $Y_i \overset{ind.}{\sim} Bernoulli(p_i)$ where $p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$ for a non-random covariate x_i.

   a) Write down the likelihood $L(\alpha, \beta; y_1, ..., y_n)$. Hint: first write down the likelihood in terms of p_i and then substitute the expression for p_i in terms of α andβ.

   b) Take the natural logarithm of the likelihood to obtain the loglikelihood $\ell(\alpha, \beta; y_1, ..., y_n)$.

   c) Find the estimating equations used to derive maximum likelihood estimators for α and β but do not attempt to solve them.

The likelihood is

$$\begin{aligned} L(\alpha, \beta; y_1, ..., y_n) &= \prod_{i=1}^n p_i^{y_i}(1-p_i)^{1-y_i} \\ &= \prod_{i=1}^n \left(\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{y_i} \left(\frac{1}{1+e^{\alpha+\beta x_i}}\right)^{1-y_i} \\ &= \prod_{i=1}^n \frac{e^{y_i(\alpha+\beta x_i)}}{1+e^{\alpha+\beta x_i}}. \end{aligned}$$

The loglikelihood is

$$\begin{aligned} \ell(\alpha, \beta; y_1, ..., y_n) &= \sum_{i=1}^n \log\left[\frac{e^{y_i(\alpha+\beta x_i)}}{1+e^{\alpha+\beta x_i}}\right] \\ &= \sum_{i=1}^n [y_i(\alpha+\beta x_i) - \log(1+e^{\alpha+\beta x_i})] \end{aligned}$$

Take partial derivatives with respect to α and β:

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n}(y_i - p_i)$$

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n}(y_i x_i - p_i x_i)$$

The estimating equations are:

$$0 = \sum_{i=1}^{n}(y_i - p_i)$$

$$0 = \sum_{i=1}^{n}(y_i x_i - p_i x_i)$$

4. (6 pts.) Define the median $x_{0.50}$ of a distribution with CDF $F(x)$ and pdf $f(x)$. Let $k = \lfloor 0.5(n+1) \rfloor$ and consider the order statistic $X_{(k)}$.

   a) It can be shown that $X_{(k)} \dot\sim N(\mu = x_{0.50}, \sigma^2 = \frac{1}{4nf(x_{0.50})^2})$ for large n. Derive a large sample approximate confidence interval formula for the median $x_{0.50}$.

   b) Why might your CI in part a) be difficult to implement in practice? Can you suggest an alternative approach? You can just describe this approach in words, a formula is not necessary.

$$(x_{(k)} + z_{\alpha/2}/\sqrt{4nf(x_{0.50})^2},\ x_{(k)} + z_{1-\alpha/2}/\sqrt{4nf(x_{0.50})^2})$$

Since we do not know $f(x_{0.50})$ we cannot actually compute this CI given data. We might use a parametric approach, that is, assume some form of $f(x)$, like a normal distribution. In that case we might use the CI for the mean. Otherwise, we could consider a "nonparametric" CI for $x_{0.50}$ in which we find order statistics $(X_{(s)}, X_{(t)})$ such that the binomial probability between these values is about $1 - \alpha$.

5. (8 pts.) Suppose X_1 and X_2 are independent Poisson random variables. Recall the Poisson pmf

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!},\, x = 0, 1, 2, ...$$

where $\lambda > 0$. Suppose we will test $H_0 : \lambda = 2$ versus $H_1 : \lambda = 1$, so $\lambda \in \{1, 2\}$ (there are only these two possible values). We reject H_0 is

$$\frac{p(x_1; 2)p(x_2; 2)}{p(x_1; 1)p(x_2; 1)} \leq 0.5.$$

a) The moment generating function (MGF) of a Poisson r.v. X is

$$M_X(t) = e^{\lambda(e^t-1)}.$$

Use the MGF to show that if $Y_1 \sim Pois(\lambda_1)$ independent from $Y_2 \sim Pois(\lambda_2)$, then $Y_1 + Y_2 \sim Pois(\lambda_1 + \lambda_2)$.

b) Find the type 1 error probability of the test.

c) Find the power of the test.

$$M_{Y_1+Y_2}(t) = E(e^{t(Y_1+Y_2)}) = E(e^{tY_1}e^{tY_2}) \overset{ind.}{=} M_{Y_1}(t)M_{Y_2}(t)$$

$$= e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}.$$

Hence $Y_1 + Y_2 \sim Pois(\lambda_1 + \lambda_2)$.

We reject H_0 if $X_1 + X_2 \leq \frac{\log e^2/2}{\log 2} =: a$ and $\lfloor a \rfloor = 1$. Using the hint:

$$P(X_1 + X_2 \leq 1) = 5e^{-4} = 0.0915782.$$

Similarly, the power is

$$P(X_1 + X_2 \leq 1 | X_1 + X_2 \sim Pois(2)) = 0.406.$$

Table II
Normal Distribution

The following table presents the standard normal distribution. The probabilities tabled are

$$P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw.$$

Note that only the probabilities for $z \geq 0$ are tabled. To obtain the probabilities for $z < 0$, use the identity $\Phi(-z) = 1 - \Phi(z)$. At the bottom of the table, some useful quantiles of the standard normal distribution are displayed. The R function `normaltable.s` generates this table.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

α	0.400	0.300	0.200	0.100	0.050	0.025	0.020	0.010	0.005	0.001
z_α	0.253	0.524	0.842	1.282	1.645	1.960	2.054	2.326	2.576	3.090
$z_{\alpha/2}$	0.842	1.036	1.282	1.645	1.960	2.241	2.326	2.576	2.807	3.291

708 Tables of Distributions

Table I
Chi-Square Distribution

The following table presents selected quantiles of chi-square distribution, i.e., the values x such that

$$P(X \leq x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2}\, dw,$$

for selected degrees of freedom r. The R function `chistable.s` generates this table.

	$P(X \leq x)$							
r	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892