Statistics Practice Questions

Try to do this these questions by hand or using a calculator (no computers).

1. In a production lot of 1 million items, there are 100 defective pieces.

   (a) Write down an expression for the probability that in a random sample (without replacement) of 1000 items, there is no defective piece.

   (b) Write down an expression for the probability that in a random sample (with replacement) of 1000 items, there is no defective piece.

   (c) Do you expect the two answers above to be roughly equal numerically?

   (d) Estimate this answer numerically, by hand.

2. Information collected to compare new car models includes engine size in litres, number of cylinders, size of car (compact, mid size, full size), type of transmission (automatic or manual), and dealer cost. Classify each variable as categorical (nominal or ordinal) or numerical (discrete or continuous).

3. A company employs 2000 men and 1500 women, of whom 100 men and 50 women are in management positions. A survey is planned to study employee attitudes towards retirement contributions. Attitudes appear to differ by gender and by management status. Describe a sampling plan that would accurately reflect the attitudes for each subgroup defined by gender and management status, based on a total sample of 700 people.

4. An analyst is conducting a satisfaction survey, sampling from a list of 1000 car buyers, made up of 250 Ford buyers, 250 GM buyers, 250 Honda buyers, and 250 Toyota buyers. The analyst selects a sample of 40, by randomly sampling 10 buyers of each brand. Is this an example of a simple random sampling? Justify your answer.

5. If the sample standard deviation of the sample $\{1,2,x\}$ is 1, find x.

6. For the unmatched count example in Week 1 Lecture 1, suppose now the first group had m respondents and a total of u_1, while the second group had n respondents and a total of u_2. How can we estimate the proportion of people with racist attitudes?

7. If we could reject a null hypothesis at the 5% significance level, then is it true that we could also reject it at the 1% level? What about at the 10% level? Justify your answer.

8. (a) A study shows that 19% of unemployed men wear spectacles, and 37% of employed men wear spectacles. What is wrong with the conclusion that 'work is bad for your eyesight'?

   (b) A study shows that vegetarians on average live longer. Does it follow that a vegetarian diet is good for longevity?

   (c) A study reports that 'in the 35 years since marijuana laws stopped being enforced, the number of marijuana smokers in California has doubled every year.' Prove that this cannot possibly be the case.

9. Is this a Latin square?

1	2	3	4	5
2	5	1	3	4
3	4	5	1	2
4	1	2	5	3
5	4	3	2	1

10. At the canteen, you order from the same shop at a uniformly randomly chosen time (between opening and closing hours) on each weekday, and note down the number x_i on your order. Then does $2\bar{x}-1$ give a reasonable estimate for the total number of customers who order from the shop each day? Justify your answer.

11. In a set of data, if $\tilde{x} = 140$ and $Q_3 = 180$, what can you say about the proportion of data values that are greater than 160?

12. If you sample from a normally distributed population, then estimate the probability that a data point is classified as an outlier (according to the rule used in boxplots).

13. Assume that σ is known and n is large. For the hypothesis test $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$, find the formula for the power.

14. Suppose we set up an experiment to compare the success rates of two treatments, A and B. It is known that the treatments work differently on males and females. The samples sizes are given below.

	Male	Female
Treatment A	50	90
Treatment B	60	30

After the conclusion of the experiment, treatment A is found to have a 50% success rate *overall*, while treatment B's overall success rate is higher.

Come up with a number of successes for each cell, so that Simpson's paradox occurs.

15. (a) For the male and female body temperature data from Week 5, describe how to make a Q-Q plot.

(b) Construct any scenario involving body temperature measurements, that would benefit from a matched pairs design.

16. A Type I error is made when:

(a) The null hypothesis is rejected when it is true

(b) The null hypothesis is not rejected when it is false

(c) The null hypothesis is accepted when it is false

(d) The alternative hypothesis is rejected when it is true

(e) The alternative hypothesis is rejected when it is false

17. (a) What is the probability of a Type II error when $\alpha = 0.1$?

(b) What is the probability of a Type II error when $\alpha = 0.05$ and $n = 100$?

18. Assume that a population is normally distributed and the standard deviation is known. In order to cut the width of a two-sided confidence interval of the population mean by half, we need to increase the sample size by a factor of how many?

19. Does EWMA use all of the previous data values? Does (ordinary) moving average use all of the previous data values?

20. If an 80% confidence interval for the mean is $[7.5, 8.3]$, which of the following could be the 90% confidence interval for the mean?

(a) [7.39, 8.3]

(b) [7.8, 8.0]

(c) [7.5, 8.41]

(d) $[7.39, 8.41]$

(e) $[7.3, 8.9]$

21. Based on 11 observations, the sample variance of the speeds measured by a radar gun is calculated to be 5 $(\text{km/h})^2$. *One* of the following is the 95% two-sided confidence interval for the true variance. Even though you are not provided with a chi-squared distribution table, determine which one.

    (a) $[5.70, 20.32]$

    (b) $[2.44, 15.40]$

    (c) $[2.73, \infty)$

    (d) $[-1.03, 21.96]$

    (e) $[3.50, 6.50]$

22. Weekly postage expenses for a company have a mean of \$312 and a standard deviation of \$58. The company has allowed \$330 for postage per week in its budget. What is the approximate probability that the average weekly postage expenses for the past year exceeds the budgeted amount?

23. The nominal mean filled content of a drink is 500 ml. If the mean filled content deviates (in either direction) from this, then it must be detected. Therefore an engineer set up a hypothesis test with a significance level $\alpha = 0.01$, $H_0 : \mu = 500$ vs $H_1 : \mu \neq 500$. Suppose the filled content is distributed as $N(\mu, 4)$. The test is conducted based on hourly samples of size $n = 5$. Calculate the power of this test to detect a shift of 2 ml. What about a shift of 4 ml?

24. The distribution of the diameter of crankshafts produced for an engine is normal, with $\sigma = 0.02$ mm. A simple random sample of four crankshafts is chosen. The measured diameters are

$$200.01, 199.96, 200.01, 200.02.$$

    (a) Find the 90% two-sided, upper 90%, and lower 90% confidence intervals for μ.

    (b) Find s, and test whether it is consistent with the given σ, with $\alpha = 0.1$. You may use $\chi^2_{3,0.95} = 7.81$, $\chi^2_{3,0.05} = 0.352$.

25. From a random sample of 196 SUTD students, it is found that on average, they spend 18.0 hours on homework per week, with a sample standard deviation of 6.4 hours. Construct a 95% two-sided confidence interval, and justify why the z-distribution may be used.

26. The scores of a psychological test for university students range from 0 to 200 and follow approximately a normal distribution with $\mu = 115$ and $\sigma = 25$. You suspect that the incoming freshmores have a mean more than 115, as 25 incoming freshmores obtained a mean score of 117.5. Find the p-value.

27. A worker team has suggested an alternative changeover procedure, and hopes that it will reduce the changeover cost. The current procedure requires an average of 16.2 worker-hours with a standard deviation of 2.40 hours per changeover and is normally distributed. The manufacturer tests the new procedure on the 16 new changeovers which can be regarded as a random sample. Suppose we perform a hypothesis test with $\alpha = 0.05$. Assume that the true mean is 14.6 worker-hours. What is the power of the test in detecting this value?

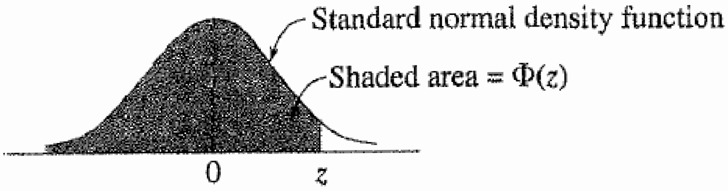
28. The pdf for a χ^2_n random variable is

$$\frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}.$$

    Recall that for a positive integer k, $\Gamma(k+1) = k!$.

    Let s^2 be the sample variance computed from a random sample of size 5 drawn from a normal distribution. Find $P(s^2 > 2\sigma^2)$.

Table A.3 Standard Normal Curve Areas $\Phi(z) = P(Z \leq z)$ (*cont.*)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	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	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	0.9278	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	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	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998