Problem set for Probability and Statistics 1 — 11 May 2026

p	0.9	0.95	0.975	0.99	0.995
$\Phi^{-1}(p)$	1.28	1.64	1.96	2.33	2.58
$\Psi_1^{-1}(p)$	3.08	6.31	12.71	31.82	63.66
$\Psi_2^{-1}(p)$	1.89	2.92	4.3	6.96	9.92
$\Psi_4^{-1}(p)$	1.53	2.13	2.78	3.75	4.6
$\Psi_8^{-1}(p)$	1.4	1.86	2.31	2.9	3.36
$\Psi_{20}^{-1}(p)$	1.33	1.72	2.09	2.53	2.85
$\Psi_{30}^{-1}(p)$	1.31	1.7	2.04	2.46	2.75

Confidence intervals

1. We take a sample of size n and use it to define a 99% confidence interval $[\hat{\theta}^-, \hat{\theta}^+]$ for a parameter θ. Which of the following are true?

(a) The distribution of θ is such that it lies in $[\hat{\theta}^-, \hat{\theta}^+]$ with probability 0.99.

(b) On average, 99% of the sample data lies within $[\hat{\theta}^-, \hat{\theta}^+]$.

(c) If we were to take a second sample of size n and estimate θ based on it, the estimator $\hat{\theta}$ lies in $[\hat{\theta}^-, \hat{\theta}^+]$ with probability 0.99.

(d) If we were to take N samples of size n and define 99% confidence intervals $[\hat{\theta}_1^-, \hat{\theta}_1^+], \dots, [\hat{\theta}_N^-, \hat{\theta}_N^+]$, 99% of these intervals on average contain θ.

2. We have one measurement $X \sim \mathcal{N}(\mu, 1)$.

(a) Find a 90% confidence interval for μ.

(b) Instead of one measurement, we perform 9 (independent) measurements. What will be the confidence interval for μ now?

(c) Let X still have mean μ and variance 1, but it is no longer necessarily normal. What changes?

(d) Let X still have mean μ but we are given no information about its variance or distribution. What can we say then?

3. This time we are sampling from $\mathcal{N}(\mu, \sigma^2)$: we do not know μ or σ, so the parameter is (μ, σ). The sampled values are 8.32, 10.92, 10.52, 9.52, 10.82.

(a) Calculate the sample mean and sample variance (with the Bessel correction).

(b) If we believed that the computed sample variance is the true value of σ^2, find a 98% confidence interval for μ.

(c) Find a 98% confidence interval for μ using Student's t-distribution.

4. We model the number of emails per day using a Poisson distribution $\mathsf{Poi}(\lambda)$. In the first week of May, we received 34, 36, 29, 31, 30 emails. Find a 95% confidence interval for λ using Student's t-distribution. (Although the Poisson distribution is not normal, for sufficiently high values of λ, it is very similar to normal, so the method will have reliability close to 95%.)

Hypothesis testing

5. We have one measurement $X \sim \mathcal{N}(\mu, 1)$. We want to test the hypothesis H_0: $\mu = 5$ at a significance level $\alpha = 20\%$.

(a) What critical region will we choose – the set of measurements where we reject the hypothesis?

(b) Instead of one measurement, we take n independent measurements. What will be the critical region for $\bar{X}_n$?

(c) If in fact $\mu = 4$ and we have $n = 10$ measurements, what is the probability that we do not reject the hypothesis, given that $\Phi(\sqrt{10} - 1.28) = 0.9701, \Phi(\sqrt{10} + 1.28) = 0.9999$?

6. * You want to determine whether or not more than 75% of citizens of country X would vote for candidate Y for President in 2028. In a random poll sampling of $n = 137$ citizens of X, we collected responses $x_1, \ldots, x_n$ (each either "yes" or "no" which we encode as 1 and 0, respectively). We observe 131 "yes" responses: $\sum_{i=1}^n x_i = 131$. Perform a hypothesis test at a significance level 5% and state your conclusion. (See the hints at the end for help.)

For problem 6, you should get an inequality relating the sum of binomial coefficients and a product of a number of terms. You can bound each of the terms involving the binomial coefficients by the product of the largest coefficient and the multiplier on the smallest coefficient and then with the identity $\binom{n}{k} \leq (\frac{en}{k})^k$. You may also use the inequalities $1.5 \leq \log_2(3) \leq 1.6$.
