STAT 515 sp 2024 Exam II

Karl Gregory

- Do not open this exam until told to do so.
- You may have one handwritten sheet of notes out during the exam.
- You have 75 minutes to work on this exam.
- You may NOT use any kind of calculator.
- If you are unsure of what a question is asking for, do not hesitate to ask me for clarification.
- *Good luck, and may the odds be ever in your favor!*

$X \sim$		$\mathcal{X}$	$\mathbb{E}X$	$\text{Var}(X)$
Poisson(λ)	$P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$	$x = 0, 1, 2, \ldots$	λ	λ
Exponential(λ)	$P(X \leq x) = 1 - e^{-x\lambda}$	$x > 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$

$$\hat{p}_n \pm z_{\alpha/2} \cdot \sqrt{\hat{p}_n(1 - \hat{p}_n)/n} \qquad \bar{X}_n \pm t_{n-1,\alpha/2} \cdot S_n/\sqrt{n}$$

$$Z_{\text{test}} = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1 - p_0)/n}} \qquad T_{\text{test}} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$

1. A type of weeds grow in a farmer's field such that when he walks through his field he encounters, on average, 10 weeds every 50 yards. Suppose the weeds grow according to a Poisson process. Here comes the farmer now; he is about to walk in his field...

   (a) Give the probability that he will walk 50 yards without encountering a single weed.

   Let $X \sim \text{Poisson}(\lambda = 10)$.

   $$P(X=0) = \frac{e^{-10}(10)^0}{0!} = e^{-10}.$$

   (b) Give the probability that he encounters at least one weed in the first 50 yards.

   $$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-10}.$$

   (c) Give an expression for the probability that he finds at least 10 weeds in the first 50 yards. You do not need to evaluate your expression.

   $$P(X \geq 10) = 1 - P(X \leq 9) = 1 - \sum_{x=0}^{9} \frac{e^{-10}(10)^x}{x!}.$$

   (d) Give the expected number of weeds he will encounter if he walks 300 yards in his field.

   Let Y = # weeds in 300 yds. Then $Y \sim \text{Poisson}(\lambda = 60)$,

   since $\frac{300}{50} = 6$ and $6 \cdot 10 = 60$.

   Now $\mathbb{E}\, Y = 60$, which is the answer.

   (e) Suppose he measures the distance from the first weed he encounters to the next weed he encounters. What probability distribution does this random variable have?

   This random variable has an exponential distribution.

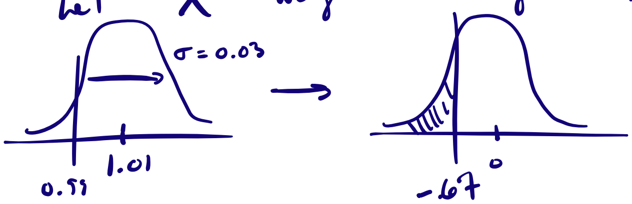
   If W = distance walked until next weed, then

   $$W \sim \text{Exponential}(\lambda = 10).$$

2. A certain purveyor's monster-size bags of potato chips have weights following a Normal distribution with mean 1.01 lbs and standard deviation 0.03.

(a) Give the probability that a randomly sampled bag will weigh less than 0.99 lbs.

Let X = weight of bag. Then $X \sim N(\mu = 1.01, \sigma^2 = (0.03)^2)$



$$\frac{0.99-1.01}{0.03} = -\frac{0.02}{0.03} = -.67$$

$$P(X < 0.99) = P(Z < -.67) = 0.5 - 0.2486 = 0.2514$$

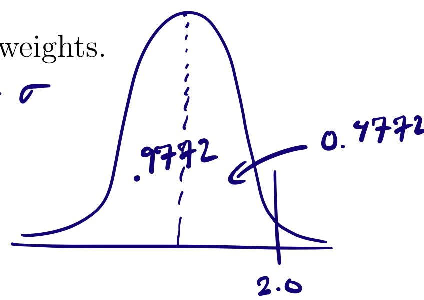
(b) Give the probability that a randemly sample big will weight more than 0.99 lbs.

This is the complement, so we have

$$P(X > 0.99) = 1 - P(X < 0.99) = 0.7486$$

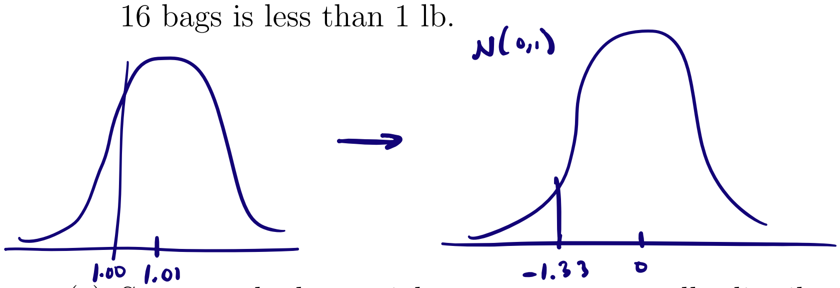
(c) Give the 97.72 percentile of bag weights.

$$Z = \frac{X-\mu}{\sigma} \iff X = \mu + Z\sigma$$



$$q_{.9772} = 1.01 + 2(0.03) = 1.07$$

(d) Suppose you take a random sample of 16 bags. Give the probability that the average weight of the 16 bags is less than 1 lb.



$$\frac{1.00-1.01}{0.03/\sqrt{16}} = \frac{-0.1}{.03/4} = -\frac{0.4}{0.3} = -1.33$$

$$P(\bar{X}_n < 1.0) = P(Z < -1.33) = 0.5 - .4082$$
$$= 0.0918$$

(e) Suppose the bag weights were *not* normally distributed. What is required for the mean weights of samples of bags to have approximately a normal distribution?

We need a large sample size, say $n \geq 30$.

3. A blood donation advocate is interested in the proportion of college freshmen who have, prior to starting college, donated blood. She draws a sample of 100 freshmen, among whom she finds that 15 donated blood before starting college.

(a) Give an expression for a 95% confidence interval for the true proportion of college freshmen who have donated blood prior to starting college. You do not need to evaluate the bounds of the interval.

$\hat{p}_n = 0.15$

C.I. is $\quad 0.15 \pm \underbrace{1.96}_{z_{\frac{0.05}{2}}} \sqrt{\frac{0.15(1-0.15)}{100}}$

(b) The researcher wishes to know whether the percentage of college freshmen who have given blood before starting college is less than 20%. Formulate a null and an alternate hypothesis which correspond to her research question.

$H_0: p \geq 0.20 \quad$ vs $\quad H_1: p < 0.20$

(c) Compute the value of the test statistic (it is possible, with some arithmetic, do to this without a calculator) for testing your hypotheses in the previous part.

$$Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.15 - 0.20}{\sqrt{\frac{.20(.80)}{100}}} = \frac{-0.05\,(10)}{\sqrt{.16}} = -\frac{.5}{.4} = -1.25.$$

(d) Does the researcher reject H_0 at the $\alpha = 0.05$ significance level?

We reject H_0 if $Z_{test} < -z_{0.05} = -1.645$

We fail to reject H_0 since $Z_{test} = -1.25$.

(e) Suppose the researcher plans to do a larger study; she wishes in the end to be able to construct a confidence interval for the true proportion of freshmen who have given blood before starting college with a margin of error no greater than 3 percentage points. Give an expression for a guess of the sample size needed for her to achieve this.

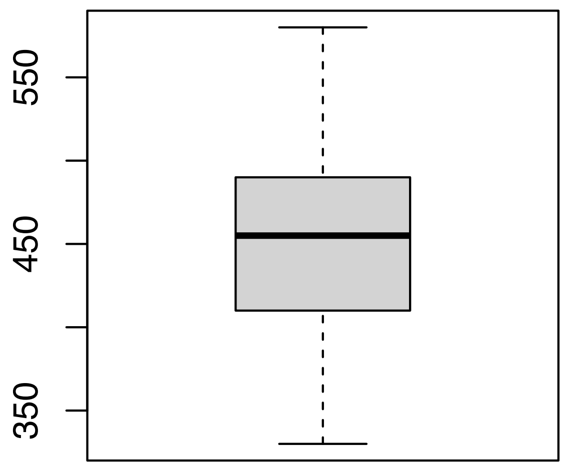
Take $\quad n \geq \left(\frac{1.96 \sqrt{.15(.85)}}{.03} \right)^2.$

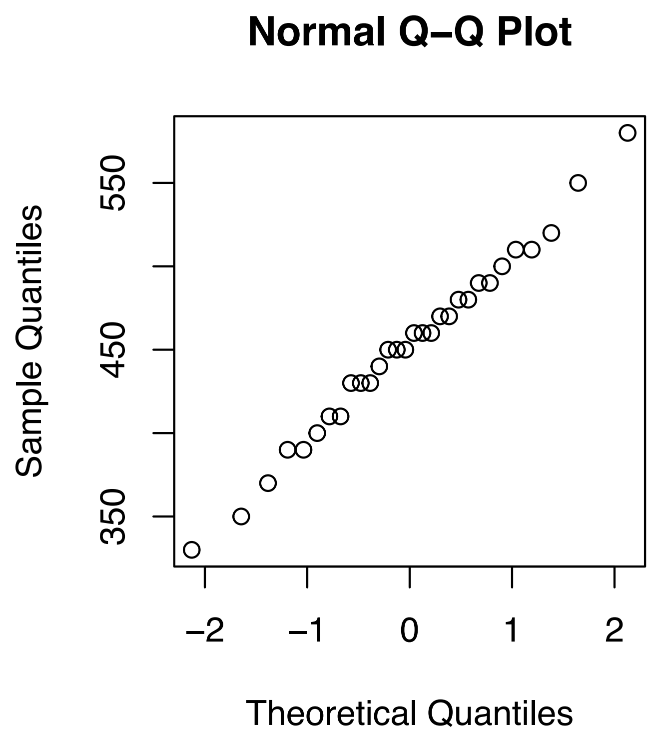
4. Municipal water quality investigators take 30 water samples and measure in each one the level of organic pollutants, obtaining the measurements below:

330	390	410	430	450	460	470	480	500	520
350	390	410	430	450	460	470	490	510	550
370	400	430	440	450	460	480	490	510	580

$n=30$

The average of all the measurements is 452 and the standard deviation is 56.47. Boxplots and a Normal quantile-quantile plot are shown.





(a) Give an expression for a 95% confidence interval for the true pollutant level in the water. You do not have to evaluate the endpoints of your interval.

Use $\bar{X}_n \pm t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}$. This becomes

$$452 \pm \underbrace{2.0452}_{t_{29,\,0.025}} \frac{56.47}{\sqrt{30}}$$

(b) Municipal authorities will take action if it is determined that the true pollutant level exceeds 430. Write down the relevant null and alternate hypotheses.

$H_0: \mu \leq 430$ vs $H_1: \mu > 430$

(c) Give an expression for the test statistic for testing the hypotheses in the previous part. You do not have to evaluate your expression.

$$T_{test} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{452 - 430}{56.47/\sqrt{30}}$$

(d) The value of the test statistic is 2.134. State your decision about the hypotheses in part (b) at the $\alpha = 0.05$ as well as at the $\alpha = 0.01$ significance level.

The rejection thresholds are $t_{29,0.05} = 1.699$

and $t_{29,0.01} = 2.462$.

We reject H_0 at $\alpha = 0.05$ but not at $\alpha = 0.01$.

(e) Explain the purpose of the Normal quantile-quantile plot.

This helps us see whether the sample was drawn from a Normal distribution.