Midterm 1 Review

https://dzwang91.github.io/stat324/

Announcements

- Midterm 1 will take place this Thursday at this room.
- Midterm 1 will cover all materials through QQ plot and CLT.
- You are allowed to take a **two-side** note and calculator. Statistical tables will be provided.
- 10 problems in total.

Outline

What you should know

- Sample mean
- Sample standard deviation
- Median
- Q_1, Q_3
- Range
- IQR
- Histogram: you need understand how to extract information from the histogram.

Outline

What you should know

- Random process, outcome, sample space, event
- P(E)=sum of probabilities of outcomes in E
- $0 \leq P(E) \leq 1$
- P(not E)=1-P(E)
- A and B are independent if occurrence of one doesn't change the probability of the other, then P(A and B)=P(A)P(B).

Outline

What you should know

- Probability mass function (pmf): p(x)=p(X=x)
- Mean: $\mu = \mathbb{E}(X) = \sum_x xp(x)$.
- Mean properties:

$$\mathbb{E}(c) = c, \mathbb{E}(cX) = c\mathbb{E}(X),$$

$$\mathbb{E}(X + c) = \mathbb{E}(X) + c, \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

- Variance: $\sigma^2 = \mathbb{E}[(X - \mu)^2] = \sum_x (x - \mu)^2 p(x)$.
- Variance properties:

$$VAR(c) = 0, VAR(cX) = c^2 VAR(X), VAR(X + c) = VAR(X).$$

For independent X and Y,

$$VAR(X + Y) = VAR(X) + VAR(Y).$$

What you should know

- Bernoulli RV: $\mathbb{P}(Y = 1) = \pi, \mathbb{P}(Y = 0) = 1 - \pi$.
- $\mu = \pi, \sigma^2 = \pi(1 - \pi)$.
- Binomial RV: $X \sim Bin(n, \pi)$ is number of successes in n independent Bernoulli trials, each with P(success)$=\pi$.
- $\mathbb{P}(X = x) = \frac{n!}{x!(n-x)!}\pi^x(1 - \pi)^{n-x}$ for $x = 0, ..., n$.
- $\mu = n\pi$, $\sigma^2 = n\pi(1 - \pi)$.

Outline

What you should know

- $\mathbb{P}(a \leq X \leq b)$: area under $f(x)$ between a and b, where $f(x)$ is the pdf.
- cumulative distribution function (cdf): $F(x) = \mathbb{P}(X \leq x)$.
- Normal distribution: $N(\mu, \sigma^2)$, μ is the mean and σ^2 is the variance.
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$.
- If $Z \sim N(0, 1^2)$, then $X = Z\sigma + \mu \sim N(\mu, \sigma^2)$.
- $P(X < x) = P(Z = \frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma})$.
- You should know how to use Z table.

Outline

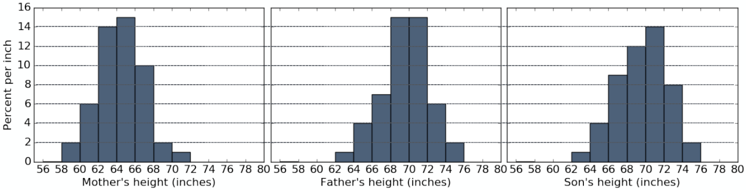
What you should know

- Simple random sample (SRS)
- $X_1, ..., X_n$ are iid with mean μ and variance σ^2, then $\mathbb{E}(\bar{X}) = \mu$, $VAR(\bar{X}) = \frac{\sigma^2}{n}$.
- Estimator
- Bias, standard error, MSE
- QQ plot, CLT

Outline

Problem 1

Steven measured the heights of **100 families** that each included 1 mother, 1 father, and some varying number of adult sons. The three histogram of heights below depict the distributions for all mothers, fathers and adult sons. **All bars are 2 inches wide. All bars' heights are integers.** The heights of all people in the data set are included in the histograms.



Problem 1 continued

- The **number** of sons that are at least 70 inches tall.

Problem 1 continued

- The **number** of sons that are at least 70 inches tall.
  Unknown

Problem 1 continued

- The **number** of sons that are at least 70 inches tall.
  Unknown
- If the sons' histogram were redrawn, replacing the two bins from 72-to-74 and from 74-to-76 with one bin from 72-to-76, what would be the height of its bar? If it's impossible to tell, write **unknown**.

Problem 1 continued

- The **number** of sons that are at least 70 inches tall.
  Unknown
- If the sons' histogram were redrawn, replacing the two bins from 72-to-74 and from 74-to-76 with one bin from 72-to-76, what would be the height of its bar? If it's impossible to tell, write **unknown**.

$$0.08 \times 2 + 0.02 \times 2 = 0.2$$

$$0.2/4 = 0.05 = 5\%$$

Problem 2

Three fans was randomly selected at the football game in which Madison was playing Minnesota in November 2018. Since this is a home game, we assume 90% of the fans attending the game are Madison fans, while 10% are Minnesota fans. Let X be the number of Madison fans selected, answer following questions:

- List a table of **pmf** of X.

Problem 2

Three fans was randomly selected at the football game in which Madison was playing Minnesota in November 2018. Since this is a home game, we assume 90% of the fans attending the game are Madison fans, while 10% are Minnesota fans. Let X be the number of Madison fans selected, answer following questions:

- List a table of **pmf** of X.

X	pmf
0	0.001
1	0.027
2	0.243
3	0.729

- Calculate the **expectation** and **variance** of X.

Problem 2

Three fans was randomly selected at the football game in which Madison was playing Minnesota in November 2018. Since this is a home game, we assume 90% of the fans attending the game are Madison fans, while 10% are Minnesota fans. Let X be the number of Madison fans selected, answer following questions:

- List a table of **pmf** of X.

X	pmf
0	0.001
1	0.027
2	0.243
3	0.729

- Calculate the **expectation** and **variance** of X.

$$E(X) = 3 \times 0.9 = 2.7$$

$$VAR(X) = 3 \times 0.9 \times 0.1 = 0.27$$

Problem 2 continued

- If we further assume two persons are friends only if they are same team's fans, then what's the probability that these selected three people are friends with each other?

Problem 2 continued

- If we further assume two persons are friends only if they are same team's fans, then what's the probability that these selected three people are friends with each other?

$$P(X = 0 \text{ or } X = 3) = P(X = 0) + P(X = 3) = 0.73$$

Problem 3

A factory manager has been working in the factory for many years. Sometimes the old machine will break down in the middle of production which causes some problems for timely delivery. Assume that the time of an old machine stays in operation before a maintenance is needed is normally distributed with **mean 500 hours and standard deviation of 100 hours**.

- What is the probability that an old machine could operate at least 600 hours before the next maintenance is arranged?

Problem 3

A factory manager has been working in the factory for many years. Sometimes the old machine will break down in the middle of production which causes some problems for timely delivery. Assume that the time of an old machine stays in operation before a maintenance is needed is normally distributed with **mean 500 hours and standard deviation of 100 hours**.

- What is the probability that an old machine could operate at least 600 hours before the next maintenance is arranged?

$$P(X \geq 600) = P(Z \geq \frac{600 - 500}{100}) = P(Z \geq 1) = 1 - 0.84 = 0.16$$

Problem 3

- If 10 machines are randomly selected, what is the probability that at least 2 would be able to operate at least 600 hours? Assume that the machines work independently.

Problem 3

- If 10 machines are randomly selected, what is the probability that at least 2 would be able to operate at least 600 hours? Assume that the machines work independently.
  Let B=the number of machines which are able to operate at least 600 hours, then $B \sim Bin(10, 0.16)$, then

  $$P(B \geq 2) = 1 - P(B = 0) - P(B = 1) = 0.49.$$

- If 2 machines are randomly selected, what is the probability that one would be able to operate for at least 600 hours while the other could only operate at most 600 hours?

Problem 3

- If 10 machines are randomly selected, what is the probability that at least 2 would be able to operate at least 600 hours? Assume that the machines work independently.
  Let B=the number of machines which are able to operate at least 600 hours, then $B \sim Bin(10, 0.16)$, then

$$P(B \geq 2) = 1 - P(B = 0) - P(B = 1) = 0.49.$$

- If 2 machines are randomly selected, what is the probability that one would be able to operate for at least 600 hours while the other could only operate at most 600 hours?

$$2 \times 0.16 \times 0.84 = 0.2688$$

Problem 4

We have a fair four-sided die with sides numbered 1-4. If we roll it twice and got at least one 4, what is the probability of rolling a 3 for the first time and a 4 for the second time?

Problem 4

We have a fair four-sided die with sides numbered 1-4. If we roll it twice and got at least one 4, what is the probability of rolling a 3 for the first time and a 4 for the second time?

There are 7 outcomes that include at least one 4. So the probability of rolling a 3 for the first time and a 4 for the second time is $1/7$.

Problem 5

Let θ be a patient's true blood pressure (in mm Hg). Suppose the patient's doctor has two devices for measuring blood pressure that are subject to random error. Let V be the blood pressure measurement taken by the first device, and let W be the blood pressure measurement taken by second device. Assume $E(V) = E(W) = \theta$ and that $Var(V) = 0.36$ while $Var(W) = 0.64$. Assume V and W are independent.

Problem 5 continued

- Suppose the doctor combines both measurements in the following way: she measures blood pressure using the first device and records V, then she measures it using the second device and records W, and records the final blood pressure measurement as $X = \frac{1}{2}(V + W)$. Calculate the **expectation** and **standard deviation** of X.

Problem 5 continued

- Suppose the doctor combines both measurements in the following way: she measures blood pressure using the first device and records V, then she measures it using the second device and records W, and records the final blood pressure measurement as $X = \frac{1}{2}(V + W)$. Calculate the **expectation** and **standard deviation** of X.

$$E(X) = E\left(\frac{V+W}{2}\right) = \frac{1}{2}E\left(V+W\right) = \frac{1}{2}\left(E(V) + E(W)\right) = \theta$$

$$Var(X) = Var\left(\frac{V+W}{2}\right) = \frac{1}{4}Var\left(V+W\right) = \frac{1}{4}$$

Problem 5 continued

- The doctor consults a statistician, who tells her she should combine the blood pressure measurements using the formula $Y = 0.64V + 0.36W$. Calculate the **expectation** and **standard deviation** of Y.

Problem 5 continued

- The doctor consults a statistician, who tells her she should combine the blood pressure measurements using the formula $Y = 0.64V + 0.36W$. Calculate the **expectation** and **standard deviation** of Y.

$$E(Y) = E(0.64V + 0.36W) = 0.64E(V) + 0.36E(W) = \theta.$$

$$Var(Y) = Var\,(0.64V + 0.36W) = (0.64^2)Var(V) + (0.36^2)Var(W)$$

$$= (0.64^2) * 0.36 + (0.36^2) * 0.64 = 0.2304,$$

$$SD(Y) = \sqrt{0.2304} = 0.48.$$